Détection séquentielle de changements brusques des caractéristiques spectrales d'un signal numérique

Michèle Basseville, Albert Benveniste

▶ To cite this version:

Michèle Basseville, Albert Benveniste. Détection séquentielle de changements brusques des caractéristiques spectrales d'un signal numérique. [Research Report] RR-0129, INRIA. 1982. inria-00076431

HAL Id: inria-00076431
https://inria.hal.science/inria-00076431

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IRIA

CENTRE DE RENNES

IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél. 954 90 20

Rapports de Recherche

N° 129

DÉTECTION SÉQUENTIELLE DE CHANGEMENTS BRUSQUES DES CARACTÉRISTIQUES SPECTRALES D'UN SIGNAL NUMÉRIQUE

Michèle BASSEVILLE
Albert BENVENISTE

Avril 1982

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: 954 90 20

Rapports de Recherche

N° 129

DÉTECTION SÉQUENTIELLE DE CHANGEMENTS BRUSQUES DES CARACTÉRISTIQUES SPECTRALES D'UN SIGNAL NUMÉRIQUE

Michèle BASSEVILLE
Albert BENVENISTE

Avril 1982

DÉTECTION SÉQUENTIELLE DE CHANGEMENTS BRUSQUES DES CARACTÉRISTIQUES SPECTRALES D'UN SIGNAL NUMÉRIQUE

Michèle BASSEVILLE (+)

Albert BENVENISTE (++)

Publication Interne n° 161 - Mars 1982 - 90 pages

RESUME

On étudie le problème de la détection séquentielle de ruptures du comportement spectral de signaux numériques, tel qu'il se pose par exemple lors de la segmentation séquentielle de signaux non-stationnaires (parole, EEG, ECG, signaux géophysiques...). On montre les limitations des approches classiques et on propose de nouveaux algorithmes plus puissants, dont les propriétés sont étudiées tant par simulations que d'un point de vue théorique.

ABSTRACT

The problem of the sequential detection of abrupt changes in the spectral behavior of a digital signal is adressed. Some new algorithms are presented and compared to some more classical ones, both via a simulation study and from a theoritical point of view.

(+) IRISA/CNRS (++) IRISA/INRIA

Secrétariat de Rédaction : Melle F. MOINET - IRISA - Lab. Informatique
Université de Rennes - Campus de Beaulieu
35042 RENNES

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L.A. 227)
UNIVERSITÉ DE RENNES 1 I.N.S.A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

SEQUENTIAL DETECTION OF ABRUPT CHANGES IN SPECTRAL CHARACTERISTICS OF DIGITAL SIGNALS

Michèle BASSEVILLE , Albert BENVENISTE

PAPIER RECUPERE ET RECYCLE

I. INTRODUCTION

1) Segmentation and adaptive analysis

These last ten years, there has been a growing interest in nonstationary modelling and processing of digital signals, coming from various domains of application such as speech processing, image processing, automatic analysis of biomedical signals : electroencephalograms and electrocardiograms, digital transmission systems, geophysics, underwater acoustics... In this framework, the problem of segmentation between "homogeneous" parts of the signal (or detection of abrupt changes in the signal) is a key point which arises more or less explicity. Either the adaptive segmentation of the signal is interesting in itself or the information provided by the detector is used in order to update or reinitialize some filter estimates or gains for adaptive identification or filtering. Actually, two main types of problems can be distinguished :

i) segmentation of a signal, the true model of which is not known, and where the model which is used for "change" or jump detection is simply a tool for localizing the boundaries. (It should be emphasized that, as the difficulty of the change detection problem increases with the complexity of the model and the number of paremeters which can possibly "jump", a hierarchical approach may be of interest. This point will be further investigated).

ii) segmentation of a signal which is approximately represented by a large amount of models : the analysis is then of an artificial intelligence type, the changes may be not really abrupt, and are recognized according to a kind of dynamic programming using simplifying heuristics (speech processing, EEG and ECG signal processing...).

Various techniques are available, among which can be found :

a) the sliding block analysis : the signal is presegmented in not necessarily disjoint intervals or blocks, and the identification or filtering is achieved inside each block ; slow changes or transitions can be detected ; see [16] for example ;

b) the partitioning or multiple model approach, which assumes that the signal is described by one of a finite number of possible models and allows switchings among the models ; the adaptive analysis includes the adaptive (sequential) choice of the model ; see [37] for a survey and unified treatment of the available algorithms ;

c) the use of filtering techniques based upon time-varying models, such as ARMA models with time-varying AR coefficients (see [7]), or time-varying AR and MA coefficients (see [14] and [12]) ;

d) segmentation algorithms, which allow two kinds of adaptive analysis : either nonsequential filtering or identification inside each of the intervals which have been so determined (this step is necessary when the model which has been used for the "change" detection is too simple to describe or classify the corresponding segment), or sequential filtering using, after each detection of a change, a compensation scheme for the filter estimates and gains according to the informations provided by the detector ; see [4], [30], [39a], [39b] for example.

Techniques <u>a</u> et <u>b</u> are appropriate for type of problem ii), and techniques <u>c</u> and <u>d</u> for problem i). In the sequel, we shall consider only techniques of type <u>d</u>.

2) Two basic problems in the detection of abrupt changes in digital signals

a) The problems

The analysis of the behavior of many real signals shows that most of the abrupt changes that occur are either changes in mean, as it is the case for edge detection in digital pictures [2a] or for some geophysical signals [4] ; or changes in spectral characteristics as for speech ([10], [12]) and EEG ([6], [16]) processing. In systems and control design applications, where more complex models are involved, other change problems than these two basic ones may arise ; see for example the survey papers [38] and [3].

Actually, in signal processing applications, the importance and usefullness of AR and ARMA models for spectral analysis have been demonstrated for various signals (see [16], [26] for example). Some statistical theoretical motivations for the use of AR models can be found in [9]. On the other hand, when the signal is non-zero mean, as it is the case for example when the mean-level has an obvious linear (time domain) variation, the use and interpretation in spectral domain of such techniques are rather ambiguous, and a first processing for removing this mean level is often necessary. If the variations of this mean are themselves non-stationary, as it is the case for some geophysical signals (see [4]), the detection of abrupt changes in this linear behavior is of interest. Thus, the purpose of this paper will be restricted to the sequential detection of abrupt changes in spectral characteristics of zero-mean digital signals.

b) One general approach

The investigation of the literature concerning the various statistical jump or change detection techniques which are used in Signal Processing, Statistics, and Automatic Control, shows that the detection of jumps in mean is of crucial importance. See the survey papers ([38], [3]) for example. Actually, one useful approach for the detection of abrupt changes in statistical models consists in , first, filtering the observations (Y_n) through one(or several)known or identified filter and in building a test statistics of cumulative sum type which is affected by the model change via a change in drift. See Figure n° 1a. Several types of sequential detectors of jumps in mean have been analyzed both via a theoretical and a simulation study and by application to real signals (image processing [2a], geophysical signals [4]), and so this paper will be mainly devoted to the problem of the sequential construction of some cumulative sum tests in the AR Gaussian case. The limitations of a classical test will be emphasized and some new algorithms, based upon the use of two AR models, will be presented. An important detector of jumps in mean will be found in Appendix II.

II. ANALYTICAL DERIVATION OF THE TESTS

The problem which is addressed is the following one : suppose the observed scalar signal (y_n) is described by an auto-regressive (AR) model, the parameters of which may jump at some unknown time, i.e. :

$$(1) \quad \begin{cases} y_n = \sum_{i=1}^{p} a_i^{(n)} \, y_{n-i} + \varepsilon_n \\ \\ \operatorname{var}(\varepsilon_n) = \sigma_n^2 \end{cases}$$

where :

$$(2) \quad \begin{cases} \begin{cases} a_i^{(n)} = a_i^o & (1 \leq i \leq p) \\ \sigma_n^2 = \sigma_o^2 \end{cases} & \text{for } n < \theta \\ \text{and} & \\ \begin{cases} a_i^{(n)} = a_i^1 & (1 \leq i \leq p) \\ \sigma_n^2 = \sigma_1^2 \end{cases} & \text{for } n \geq \theta \end{cases}$$

and $(\varepsilon_n)_n$ is a white noise sequence.

As usual for the sequential detection of jumps or changes ([3]), the parameters $(a_i^o \ (1 \leq i \leq p), \sigma_o^2)$ before change are assumed to be known, or, if not, are identified with the aid of a convenient filter. On the contrary, the characteristics of the signal after the jump $(a_i^1 \ (1 \leq i \leq p), \sigma_1^2)$ are unknown.

This model ((1) , (2)) for changes in AR models may seem to be rather limitative with respect to the more general case where the order p may also change. But it will be shown that using some special type of identifying AR filters may prevent from too much drastic limitation.

The problems consist in the sequential determination (possibly with a "forfaitary" delay) of :

- possibly, the characteristics of the signal before change ;
- the change time θ ;
- the "magnitude" of the change, i.e. a measure of distance between the identified model before change and the model corresponding to the estimates given by the filter at the detection time.

The detection of the change has to be achieved in such a way that the standard tradeoff between false alarms and delay to detection is solved at best, and that the dissymetry of the problem is weakened (it is wellknown that, im many cases, it is far more easy to detect a change from model A to model B that the contrary).

Two types of algorithms will be described, which use either one or two autoregressive models.

1) Using one AR model

As it has been outlined in the Introduction, one useful approach for the detection of abrupt changes in spectral characteristics consists in filtering the observed signal (y_n) through a known or identified AR filter, and in looking for changes in the residual signal of "innovations" (e_n). If a bias on the model is looked for, a jump in the mean of e_n can be detected ; if jumps in other parameters of the model are expected, looking for jumps in the mean of e_n^2 may be of interest. Actually, the use of cusum techniques based upon the innovations (or one-step prediction errors) e_n or the squared innovations e_n^2 is a standard approach for change detection in AR models. (Let us notice however that one attempt has been made to detect changes in a complex linear system by using Hinkley's cumulative sum test fot detection jumps in the mean of $e_n e_{n-1}$, in an adaptive filtering framework. See [30]). Such a technique, based upon e_n^2, has been used in at least three references ([21], [8], [32]) and is based upon the simple fact that, before the change $(n < \theta)$: $E(e_n^2) = \sigma_o^2$, and thus : $E\left(\frac{e_n^2}{\sigma_o^2} - 1\right) = 0$. More generally, in the vector case, $e'_n \sum_o^{-1} e_n$ is asymptotically distributed as a χ^2, and this property has been used by R.H. JONES et al [21], who detected changes in vector EEG signals with the aid of a moving-average statistics :

$\sum_{k=n-q}^{n} e'_k \sum_o^{-1} e_k$. Some years later, L.I. BORODKIN and V.V. MOTTL' [8] applied the central limit theorem and used the test statistics :

$$(3) \qquad z_n = \frac{1}{\sqrt{2n}} \sum_{k=1}^{n} \left(\frac{e_k^2}{\sigma_o^2} - 1\right)$$

which is (for $n < \theta$) asymptotically governed by the Gaussian law $N(0,1)$ and which has (for $n > \theta$) a monotonically increasing (or decreasing) mean. When θ is large, they update z_n to zero every T steps, where T is fixed a priori as a function of the length of the periods during which the observed signal (y_n) is stationary.

It will be shown in section III that this duration T is of crucial importance, because the zero-mean behavior of z_n before change does not prevent from false alarms. One possibility for overcoming this drawback could be the use of Shiryaev's or Hinkley's stopping-time which improve the performances of the detection (see [2b], [4]), but the translatinn of any change of some AR parameters into a bias on $E(e_n^2)$ is not trivial.

Let us now describe the analytical derivation of the cumulative sum test z_n as it has been recently presented by J. SEGEN and A.C. SANDERSON [32] in a more general framework, and which can give an insight on the weakness of the test. Let the signal (y_n) be described by the <u>known</u> conditionnal densities go $(y_n \mid Y^{n-1})$ before change, and $g_1(y_n \mid Y^{n-1})$ after change, where $Y^{n-1} = (y_{n-1}, y_{n-2}, \ldots, y_1)$. Let us consider the following test statistics :

$$(4) \qquad U_n = \sum_{i=1}^{n} T_i$$

where :

$$(5) \qquad T_n = \int g_n^o (y \mid Y^{n-1}) \ \text{Log } g_n^o (y \mid Y^{n-1}) \, dy - \text{Log } g_n^o (y_n \mid Y^{n-1}) \ ;$$

this statistics has a zero drift when no change occurs ans a strictly positive drift after a change from g^o to g^1 has happened, provided that the new conditionnal law g^1 does not decrease the conditionnal entropy, i.e. :

$$(6) \qquad \int g_n^1 (y \mid Y^{n-1}) \text{ Log } g_n^1 (y \mid Y^{n-1}) \, dy \leq \int g_n^o (y \mid Y^{n-1}) \text{ Log } g_n^o (y \mid Y^{n-1}) d\bar{y}.$$

In the AR(p) gaussian case we find the previous cumulative sum :

$$(7) \qquad U_n = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{e_i^2}{\sigma_o^2} - 1\right).$$

The weakness of this test is due to the fact that the two conditional probability laws g^o and g^1 before and after the change, are compared mostly via their self-entropy, and that their mutual entropy is not at all taken into account. This point will be further investigated in the next paragraph and will give rise to a new algorithm.

Moreover, after the change, the conditional mean value of T_n is :

$$
\begin{aligned}
\mathbb{E}_{H_1}(T_n \mid Y^{n-1}) &= \int g_n^o(y \mid Y^{n-1}) \operatorname{Log} g_n^o(y \mid Y^{n-1})\, dy \\
&- \int g_n^1(y \mid Y^{n-1}) \operatorname{Log} g_n^o(y \mid Y^{n-1})\, dy \\
&= \int g_n^o(y \mid Y^{n-1}) \operatorname{Log} g_n^o(y \mid Y^{n-1})\, dy \\
&- \int g_n^1(y \mid Y^{n-1}) \operatorname{Log} g_n^1(y \mid Y^{n-1})\, dy \\
&+ \int g_n^1(y \mid Y^{n-1}) \operatorname{Log} \frac{g_n^1(y \mid Y^{n-1})}{g_n^o(y \mid Y^{n-1})}\, dy \\
&= H^{Y^{n-1}}(g_n^1) - H^{Y^{n-1}}(g_n^o) + I^{Y^{n-1}}(g_n^1 \mid g_n^o)
\end{aligned}
\tag{8}
$$

where $H^{Y^{n-1}}$ and $I^{Y^{n-1}}$ are the conditional entropy and Kullback's information, respectively. So condition (6) is obviously a <u>sufficient</u> condition for the positivity of the drift of U_n after the change. Furthermore, in the AR gaussian case, with :
$A^j = (a_1^j, \ldots, a_p^j) \quad (j = 0, 1)$:

$$
\begin{aligned}
V_n &= \int g_n^1(y \mid Y^{n-1}) \operatorname{Log} g_n^o(y \mid Y^{n-1})\, dy \; - \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{(y - A^1 Y^{n-1})^2}{2\sigma_1^2}} \operatorname{Log}\left[\frac{1}{\sigma_o\sqrt{2\pi}}\, e^{-\frac{(y - A^o Y^{n-1})^2}{2\sigma_o^2}}\right] dy \\
&= -\operatorname{Log}(\sigma_o\sqrt{2\pi}) - \frac{1}{2\sigma_o^2}\int_{-\infty}^{+\infty} (y - A^o Y^{n-1})^2\, \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{(y - A^1 Y^{n-1})^2}{2\sigma_1^2}}\, dy
\end{aligned}
$$

$$
V_n = -\operatorname{Log}(\sigma_o\sqrt{2\pi}) - \frac{1}{2\sigma_o^2}\left[\sigma_1^2 + ((A^o - A^1)\, Y^{n-1})^2\right] \tag{9}
$$

and thus :

$$(10)\qquad \begin{aligned} W_n &= \int g_n^o\,(y \mid Y^{n-1})\ \mathrm{Log}\ g_n^o\,(y \mid Y^{n-1})\ dy \\ &= -\ \mathrm{Log}\ (\sigma_o \sqrt{2\pi}) - \frac{1}{2}\ . \end{aligned}$$

Therefore : $\mathbb{E}_{H_1}\ (T_n \mid Y^{n-1}) = W_n - V_n$

$$= \frac{1}{2}\left[\frac{\sigma_1^2}{\sigma_o^2} - 1 + \frac{((A^o - A^1)\ Y^{n-1})^2}{\sigma_o^2}\right].$$

In the present case, condition (6) is nothing but : $\sigma_1 \geq \sigma_o$; but it is only a sufficient condition for the positivity of drift of T_n after the change, and it is clear from the above formula that a necessary and sufficient condition is difficult to be assessed : when $\sigma_1 < \sigma_o$, the drift of T_n can be impredictably positive or negative.

For all the different reasons which have just been mentioned, it seems that the use of only one AR model and the classical approach consisting in testing how much the sequence of innovations (e_n) is far from hypothesis "zero-mean, white noise", is not sufficient for detecting abruptly changing spectral characteristics, unless for detecting spikes in EEG signals (see [16]) or pitch in speech (see [25]) for which some convenient tests have been elaborated with the aid of only one model ; for example, in [25] D.T.L. LEE and M. MORF derived a test for detecting pitch, which is related to statistical methods for detecting outliers and is based upon the joint log-likelihood of a fixed number N of observations : $\mathrm{Log}\ p_\theta\ (y_n\ ,\ y_{n-1}\ ,\ldots,\ y_{n-N+1})$. Therefore the use of two AR models seems to be preferable ; this question will be the subject of the next paragraph.

2) Using two AR models

a) Two possible approachs

The main idea underlying the following techniques of segmentation consists in the comparison of two AR models estimated at different locations in the signal, in order to detect non-stationarities ; non-stationarity is here implicitly understood as abrupt changes in the parameters, and it will be shown that this is not a real limitation. The first attempt for deriving such an algorithm seems to have been made by G. BODENSTEIN and H. PRAETORIUS [6] for the segmentation of EEG signals, and their method has also been used by M. MATHIEU [27] for example. It proceeds as follows : an AR model is adjusted in a fixed window of size about 200 sample points, and another AR model is estimated in a moving window with the same size. When the two models are too much different from each other, the segmentation is done, and the second model becomes the reference model and so on. The distance measure between spectral models which has been used in ([6], [27]) is the mean quadratic difference between the two spectra, i.e. the euclidian distance between the correlation coefficients of the two prediction errors. The analysis of this criterion has shown that it has a high variance and has a non symetrical behavior for detecting transitions from slow waves to quick waves or the opposite case. (see [27]).

Intuitively this idea can be thought to be improved by using, as the "reference" model M_o, a "global" (and no longer local) one, i.e. by taking into account all the informations contained in the signal, and not only the first ones. See figure n°1b. This idea does not seem to have been yet explored. If the global model M'_o is estimated by a filter with slight forgetting ability, it will be slightly affected by the change, and far more precisely identified than the model M_o. Therefore, the "false" alarms should be less frequent.

Actually, this idea is more or less explicitly underlying the generalized likelihood ratio (GLR) algorithm ([39a], [39b], [4]) : the "global" model (when no change has occured) is sequentially estimated, and the "local" one (after a possible change) is estimated in all the possible windows having as end-points the possible change time and the current time n, and these models are compared via a likelihood ratio technique. More precisely, under the hypothesis of no change, the (joint) log-likelihood of the observations is, at time n, such that for each $\theta \leq n$:

$$\begin{aligned}\text{Log } p_o (y_n , \dots , y_1) &= \text{Log } p_o (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_1) \\ &+ \text{Log } p_o (y_{\theta-1} , \dots , y_1) \\ &= \text{Log } p_o (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_{\theta-p}) \\ &+ \text{Log } p_o (y_{\theta-1} , \dots , y_1)\end{aligned}$$

when the process (y_n) is assumed to be p-markovian, as it is the case for AR models. Under the hypothesis that a change occured at time $\theta \leq n$, the log-likelihood is :

$$\begin{aligned}&\text{Log } p_1 (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_1) + \text{Log } p_o (y_{\theta-1} , \dots , y_1) \\ = \; &\text{Log } p_1 (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_{\theta-p}) + \text{Log } p_o (y_{\theta-1} , \dots , y_1) ,\end{aligned}$$

using again the Markov property.

The log-likelihood ratio of these two hypothesis is therefore :

$$\begin{aligned}\Lambda (n,\theta) &= \text{Log } p_1 (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_{\theta-p}) \\ &- \text{Log } p_o (y_n , \dots , y_\theta \mid y_{\theta-1} , \dots , y_{\theta-p}) \\ &= \sum_{k=\theta}^{n} \text{Log } \frac{p_1 (y_k \mid y_{k-1} , \dots , y_{k-p})}{p_o (y_k \mid y_{k-1} , \dots , y_{k-p})}\end{aligned}$$

In the AR gaussian case :

$$\Lambda\ (n,\theta) = \frac{1}{2} \sum_{k=\theta}^{n} \left[\text{Log}\ \frac{\sigma_o^2}{\sigma_1^2} + \frac{(y_k - a_1^o\ y_{k-1} - \ldots - a_p^o\ y_{k-p})^2}{2\ \sigma_o^2} - \frac{(y_k - a_1^1\ y_{k-1} - \ldots - a_p^1 y_{k-p})^2}{2\ \sigma_1^2} \right]$$

The generalized likelihood ratio technique consists in replacing the unknown AR coefficients $a_1^o, \ldots, a_p^o$ before change and $a_1^1, \ldots, a_p^1$ after change by their maximum likelihood estimates (for fixed θ), and their replacing the unknown change time θ by its maximum likelihood estimate $\hat{\theta}_n \leq n$.

The estimation of anautoregressive model being more difficult thant that of a linear one, it is more sensible to identify the "local" AR model in a window of sufficient fixed size, i.e. to constraint $\hat{\theta}_n < n - M$. The effect upon the delay for detection, as well as the estimated change time, will be investigated later.

b) <u>Three distance measures between spectral models</u>

Using this basic idea of comparing a long term estimated AR model to a short term estimated one, we propose three different algorithms which use different distance measures between these models. The two filters we use to identify the two models both realize the lattice form implementation of approximated (or exact) least squares estimates ; one is sequential, the other uses a sliding block. With the aid of such filters, the first "natural" distance which comes up to mind is the cesptral distance which is extensively used in speech processing (see [10], [26], [40] for example).

The two other measures of comparison between the two models are based upon the conditional probability laws of the observations, as in [32], and they will be now derived in the ideal situation where these laws are known. The first one is the <u>log-likelihood ratio</u>, a measure which is justified by the GLR approach as it has been shown before ; the formula is :

$$T'_n = \text{Log} \frac{g_n^1 (y_n \mid Y^{n-1})}{g_n^o (y_n \mid Y^{n-1})} ,$$

i.e. : (11) $$T'_n = \frac{1}{2} \text{Log} \frac{\sigma_o^2}{\sigma_1^2} + \frac{(e_n^o)^2}{2 \sigma_o^2} - \frac{(e_n^1)^2}{2 \sigma_1^2}$$

in the AR gaussian case, where e_n^o and e_n^1 are the respective innovations (or prediction errors) of the models 0 (before change) and 1 (after change). A detector based upon the cumulative sum :

(12) $$U'_n = \sum_{i=1}^{n} T'_i ,$$

has already been studied by M. BAGSHAW and R.A. JOHNSON [1] in a less general framework where σ^2 is supposed to be constant, which is not very realistic. On the other hand, the log-likelihood ratio between the conditional laws of the observations in the present ideal case (known models) have been also used together with Hinkley's cumulative sum test (see Appendix III) by I.V. NIKIFOROV [29]. This point will be further investigated.

The last "distance" measure involves the cross-entropy of the two conditional probability laws, and not between the joint laws as in [22], [18].

It has been explained in the previous paragraph II 1) why this could be of interest, and further motivations can be found in the work of J.E. SHORE et al. ([34], [35], [36]). Following formulas (4) - (5), we propose to use the cumulative sum :

(13) $$U''_n = \sum_{i=1}^{n} T''_i ,$$

where :

$$(14) \quad T''_n = \int g^o_n (y \mid Y^{n-1}) \, \mathrm{Log} \, \frac{g^1_n (y \mid Y^{n-1})}{g^o_n (y \mid Y^{n-1})} \, dy \; - \; \mathrm{Log} \, \frac{g^1_n (y_n \mid Y^{n-1})}{g^o_n (y_n \mid Y^{n-1})} \; .$$

Note that the only difference between T'_n and T''_n is the correction by the conditional drift before change ; the advantage of this latter cumulative sum in real situation will be shown later.

Using (9) and (10), it comes :

$$T''_n = \mathrm{Log} \frac{\sigma_o}{\sigma_1} + \frac{1}{2} - \frac{1}{2 \sigma_1^2} [\sigma_o^2 + ((A^o - A^1) \, Y^{n-1})^2] - T'_n$$

But, at time n, we have :

$$(A^o - A^1) \, Y^{n-1} = (y_n - A^1 \, Y^{n-1}) - (y_n - A^o \, Y^{n-1})$$

$$= e^1_n - e^o_n \; .$$

Thus :

$$T''_n = \frac{1}{2 \sigma_1^2} [(\sigma_1^2 - \sigma_o^2) - (e^1_n - e^o_n)^2 - \frac{\sigma_1^2}{\sigma_o^2} (e^o_n)^2 + (e^1_n)^2]$$

or :

$$(15) \quad T''_n = \frac{\sigma_o^2}{2 \sigma_1^2} [2 \frac{e^o_n \, e^1_n}{\sigma_o^2} - (1 + \frac{\sigma_1^2}{\sigma_o^2}) \frac{(e^o_n)^2}{\sigma_0^2} + (\frac{\sigma_1^2}{\sigma_o^2} - 1)] \; .$$

The conditional mean behavior of these two cumulative sum tests U'_n and U''_n is as follows. Before the change, i.e. under the hypothesis H_o that the conditional law of the observation y_n is $g^o_n (y_n \mid Y^{n-1})$, their conditional mean values are given by :

$$\mathbb{E}_{H_o}(T'_n \mid Y^{n-1}) = \int g_n^o(y \mid Y^{n-1}) \,\mathrm{Log}\, \frac{g_n^1(y \mid Y^{n-1})}{g_n^o(y \mid Y^{n-1})}\, dy$$

(16) $$= - I^{Y^{n-1}}(g_n^o \mid g_n^1) \qquad \text{(Kullback's information)}$$

$$= -\frac{1}{2}\,\mathrm{Log}\,\frac{\sigma_1^2}{\sigma_o^2} + \frac{1}{2} - \frac{1}{2\,\sigma_1^2}\,[\sigma_o^2 + (e_n^1 - e_n^o)^2]$$

which can be seen to be strictly negative, and, from (12) :

(17) $$\mathbb{E}_{H_o}(T''_n \mid Y^{n-1}) = o.$$

After the change, i.e. under the hypothesis H_1 that the observations y_n are governed by the new conditional probability law $g_n^1(y_n \mid Y^{n-1})$, the conditional mean values of U'_n and U''_n are given by :

$$\mathbb{E}_{H_1}(T'_n \mid Y^{n-1}) = \int g_n^1(y \mid Y^{n-1}) \,\mathrm{Log}\, \frac{g_n^1(y \mid Y^{n-1})}{g_n^o(y \mid y^{n-1})}\, dy$$

(18) $$= I^{Y^{n-1}}(g_n^1 \mid g_n^o)$$

$$= \frac{1}{2}\,\mathrm{Log}\,\frac{\sigma_o^2}{\sigma_1^2} - \frac{1}{2} + \frac{1}{2\,\sigma_o^2}\,[\sigma_1^2 + (e_n^o - e_n^1)^2]$$

which is strictly positive ; and, again from (12) :

$$\mathbb{E}_{H_1}(T''_n \mid Y^{n-1}) = \int g_n^o(y \mid Y^{n-1}) \,\mathrm{Log}\, \frac{g_n^1(y \mid Y^{n-1})}{g_n^o(y \mid Y^{n-1})}\, dy - \int g_n^1(y \mid Y^{n-1}) \,\mathrm{Log}\, \frac{g_n^1(y \mid Y^{n-1})}{g_n^o(y \mid Y^{n-1})}\, dy$$

$$= - I^{Y^{n-1}}(g_n^o \mid g_n^1) - I^{Y^{n-1}}(g_n^1 \mid g_n^o)$$

(19) $$= - J^{Y^{n-1}}(g_n^o , g_n^1) \qquad \text{(Kullback's divergence)}$$

$$= 1 - \frac{1}{2}\left(\frac{\sigma_o^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_o^2}\right) - \frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_o^2}\right)(e_n^1 - e_n^o)^2$$

which is known to be strictly negative, and has the interesting property of symmetry, i.e. :

$$J^{Y^{n-1}}(g_n^o , g_n^1) = J^{Y^{n-1}}(g_n^1 , g_n^o) ,$$

although it is clear that this property of the conditional divergence does not imply the symmetry of the drifts of the susum statistics U''_n (i.e. equal drift when the change occurs from g_o to g_1, or from g_1 to g_o), since :

$$\mathbb{E}_{H_o}[J^{Y^{n-1}}(g_n^o , g_n^1)] \neq \mathbb{E}_{H_1}[J^{Y^{n-1}}(g_n^o , g_n^1)].$$

Actually, a similar idea can be found in [29] where I.V. NIKIFOROV considered a quadratic form in the parameters based upon Fisher's information, which is an approximate value of Kullback's information. But the procedure in the real situation, where the models before and after change are not known, is not indicated.

The performances of this cusum test U''_n, when coupled with a more sophisticated stopping-time, will be investigated in the two next sections, both via a simulation study and from a theoretical point of view.

3) Practical implementation of the tests when the models are not known

As for as U_n is concerned, the only AR model which is used has to be sequentially identified, and therefore at the beginning of the treatment and after each detection, a "dead" time interval (of size about 200) is needed until correct identification. The same is true when using the cusum tests U'_n and U''_n : one has to wait until the sequential filter which estimates M'_o has converged, and the sliding block filter for M_1 and the test are activated.

Finally, we shall emphasize a key point related to the practical implementation of these two cumulative sum tests U'_n and U''_n, in a sequential famework. Up to now, we have derived the two test statistics T'_n and T''_n ((9) and (12)) in a classical statistical framework of hypothesis testing between two known probability laws g^o and g^1. In the real situation, not only the two laws are unknown and thus are sequentially identified, but furthermore, until the change occurs, the two estimated laws are (ideally) equal : if the two filters are suitably chosen, they correctly identify the true underlying models M'_o and M_1 which are obviously the same before change. See Figure n°1, dotted lines. And this point is crucial for giving an explanation of the much better behavior before θ of the cusum tests U'_n and U''_n based upon two models with respect to the behavior of the cusum test U_n(8) based upon one model : these three statistics are asymptotically distributed as brownian motion (under H_o), but, because of the difference between two squared innovations $(e^1_n)^2$ and $(e^o_n)^2$ coming from (nearly) the same model, U'_n and U''_n have a diffusion coefficient (or variance) the order of magnitude of which is much lower than that of U_n. This point will be further investigated in section IV devoted to some theoretical results.

Furthermore, a more sophisticated stopping-time (see Appendix II) for detecting changes in mean can be used also in this real situation, in order to give a good estimate of the change time. See section III 3) and section IV 1).

Summary table

$$g_n^j = g_n^j\,(y \mid Y^{n-1}) \qquad (j = 0,1) : \text{conditional probability law}$$

	General case	AR gaussian case
Using one model	$U_n = \sum_{i=1}^{n} T_i$ $T_n = \int g_n^o \,\mathrm{Log}\, g_n^o \,dy - \mathrm{Log}\, g_n^o$	$U_n = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{e_i^2}{\sigma_o^2} - 1\right)$
Using two models	$U'_n = \sum_{i=1}^{n} T'_i$ $T'_n = \mathrm{Log}\, \frac{g_n^1}{g_n^o}$	$T'_n = \frac{1}{2}\,\mathrm{Log}\,\frac{\sigma_o^2}{\sigma_1^2} + \frac{(e_n^o)^2}{2\,\sigma_o^2} - \frac{(e_n^1)^2}{2\,\sigma_1^2}$
	$U''_n = \sum_{i=1}^{n} T''_i$ $T''_n = \int g_n^o \,\mathrm{Log}\, \frac{g_n^1}{g_n^o}$	$T''_n = \frac{1}{2}\left[2\,\frac{e_n^o\, e_n^1}{\sigma_1^2} - \left(1 + \frac{\sigma_o^2}{\sigma_1^2}\right)\frac{(e_n^o)^2}{\sigma_o^2} + \left(1 - \frac{\sigma_o^2}{\sigma_1^2}\right)\right]$

Notations

Conditional entropy : $H^{Y^{n-1}}(g_n^j) = -\int g_n^j \,\mathrm{Log}\, g_n^j \,dy$

" Kullback's information : $I^{Y^{n-1}}(g_n^o \mid g_n^1) = \int g_n^o \,\mathrm{Log}\, \frac{g_n^o}{g_n^1}\,dy$

" " divergence :

$$J^{Y^{n-1}}(g_n^o\,,\,g_n^1) = I^{Y^{n-1}}(g_n^o \mid g_n^1) + I^{Y^{n-1}}(g_n^1 \mid g_n^o)\,.$$

III. SIMULATION STUDY

In Appendix I will be found the two filters which have been used for identifying M'_o and M_1. Both are based upon approximated least squares algorithms in lattice form. The effect of using approximated methods together with approximated distances (such a cepstral distance computed with a finite number of coefficients) will be discussed in the sequel.

For the moment, we consider the sequential filter of Burg type for identifying M'_o (see equations (2) to (10)in Appendix I), and the autocorrelation method (see [26]) in the sliding window (of size 200) for identifying M_1.

1) Comparison between the test statistics

Let us first examine the numerical examples given by L.I. BORODKIN and V.V. MOTTL' [8] who used the cumulative sum test U_n (7). (For this test, only the sequential filter is necessary). All the figures are drawn in the following manner : the upper curve is the simulated signal, containing 2000 sample points, with changes in parameters at time 1301 (vertical line) ; the lower curve is the cusum test statistics, which is computed after a "dead" zone of length 200 (before computing cusum of squared innovations, let the filter converge!). So the evolution of the test statistics under H_0 (no change) is shown for 1100 sample points. This has to be compared with [8] where the cusum U_n is set to zero every 150 sample points if no change has been detected.

As it is has been emphasized in section II 3), our complete algorithm proceeds as follows : at the beginning, the only active filter is the sequential one which identifies the model M'_o, until some reasonable convergence is reached (typically, during 200 sample points). After that, both the second filter (sliding blok) and the test are activated. The same holds after each alarm. (Therefore, two successive changes which occur within less than 200 sample points could not be detected by this method).

The first example (in Figure n° 2a) is such that the order p of the AR model is equal to 1, the variance does not change ($\sigma_o^2 = \sigma_1^2 = 1$), and the AR coefficient jumps from 0.6 to 0.1. The statistics U_n noticeably increases after change, but is rather ill behaved before. For the same signal, Figure n° 2b shows the behavior of $-U''_n$, which is much more smooth under H_o. (The apparent delay for detection, which is high because of an excessively high threshold, can be compensated for which the aid of Hinkley's stopping time. This point will be analyzed in the sequel). The same is true for a change in variance only (from 4 to 0.25) in an AR(2) model with constant parameters 0.3 and 0.5 : compare Figures n° 3a and n° 3b. The detection with U_n is very quick (mainly because e_n^2 suddenly decreases, and $\hat{\sigma}^2$ remains the same - see (8)), but with a slightly lower threshold, the detection would have occurred much before the true change-time. Finally, Figures n° 4a and n° 4b show the same results for an AR(4) process with characteristics :

(20)

$$A^o = (0.3\ ,\ 0.5\ ,\ 0\ ,\ 0) \quad ; \quad \sigma_o^2 = 0.09$$

changing to :

$$A^1 = (0.5\ ,\ -0.3\ ,\ 0.6\ ,\ -0.5) \quad ; \quad \sigma_1^2 = 0.16.$$

In the following examples, the simulated signals will be defined by the reflection coefficients k_i^j, which are related to the autoregressive coefficients a_i^j (j = 0, 1) (see formulae (11) - (12) in Appendix I). Figures n° 5a, 5b, 5c compare the behavior of U_n, $-U''_n$, U'_n when the order p = 3, $\sigma_o^2 = \sigma_1^2 = 1$, and :

(21)

$$K^o = (k_1^o\ ,\ k_2^o\ ,\ k_3^o) = (0.7\ ,\ -0.2\ ,\ 0.06)$$

$$K^1 = (k_1^1\ ,\ k_2^1\ ,\ k_3^1) = (0.9\ ,\ -0.5\ ,\ -0.04).$$

The behavior of U'_n and U''_n in a more difficult change case is shown on Figures n° 6a and 6b, for which $\sigma_o^2 = \sigma_1^2 = 1$, and :

$$(22) \qquad \begin{aligned} K^o &= (0.9\ ,\ -0.7\ ,\ 0.2) \\ K^1 &= (0.9\ ,\ -0.5\ ,\ -0.04). \end{aligned}$$

With an extra change in variance, U'_n and U''_n have the same type of behavior. Figure n° 7 shows the case where :
$\sigma_o^2 = 10$, $\sigma_1^2 = 1$, and :

$$(23) \qquad \begin{aligned} K^o &= (0.7\ ,\ -\ 0.2\ ,\ 0.06) \\ K^1 &= (0.9\ ,\ -\ 0.7\ ,\ 0.2). \end{aligned}$$

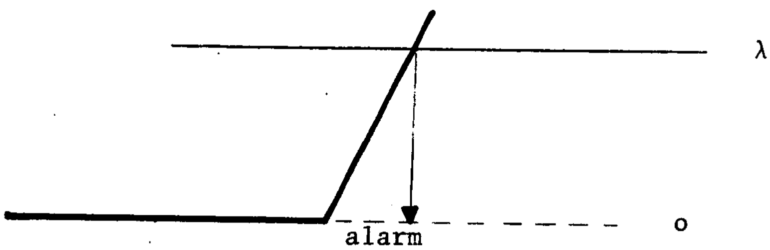
Obviously, if the size of the window used for identifying model M_1 is too small, false alarms may occur due to poor estimation of the AR coefficients. This has been observed with a size 100, both for U'_n and U''_n. So this window size has been kept to 200 samples as in [6] , [27].

Figure n° 8 shows the behavior of $-U''_n$ when using the covariance method ([26]) instead of the autocorrelation one for the sliding window, for the change of Figure n° 5. The drift under H_o seems to be slightly non zero.

Finally, let us examine the behavior of the cepstral distance d between the two identified models M'_o and M_1. In appendix I will be found the computations of the cepstral coefficients and distance starting from the estimated autoregressive coefficients. This distance seems to be noticeably sensitive to the identification method, and, when using the autocorrelation method for M_1, it is necessary to smooth the cepstral distance by computing its minimum over a small finite window (size 10 or 20). But this smoothing is no longer necessary when the covariance method is used for M_1. See Figure n° 9b for example. On Figure n° 9a are shown the variations of d^2 for the previous example (23) with K^o and K^1 exchanged. From Figure n° 10, it can be seen that in case (22), with K^o and K^1 exchanged, the cepstral distance does not seem to give a detection.

2) An estimate of the change time

Both statistics U'_n and $-U''_n$ have, for real implementation where $M'_o \equiv M_1$ until change, the following behavior :



i.e. zero drift before change, and strictly positive one after. When observing their overcrossing of a threshold, the delay for detection may be large if the threshold is too high. For reducing this delay, and obtaining a good estimate of the change time, one can apply Hinkley's cumulative sum test to detect a jump in the drift of U'_n or $-U''_n$. See [15], [2], [4]. It consists in fixing a priori a negative drift δ before jump (analog to half the minimum magnitude of jump in mean), and to observe the deviation of the new cumulative sum with respect to its minimum, i.e. to follow the scheme :

See Appendix II for further details.

From (19), it is obvious that the drift δ to be added to $-T''_n$ is related to $-\frac{1}{2} J^{Y^{n-1}} (g^1_n , g^o_n)$ (instantaneous value) ; or more exactly is the expectation of this divergence with respect to the joint probability law of the observations. (This problem will be addressed in the next section). The estimated jump time is the time at which the minimum is reached.

Some experiments have shown that $|\delta| = 0.1$ and $h = 10$ allow a delay for detection less than the size 200 for M_1, except in the non "trivial" case (22) ; and furthermore the bias of the estimated jump time is low (some units). Further results will be analyzed in the next paragraph. Figure n° 11 shows the case (23) reversed.

Finally, it should be emphasized that, although the cusum statistics U_n(8) has the same type of mean behavior as U'_n and U''_n , the sue of Hinkley's stopping-time does not seem to allow an improvement, for the two following reasons : first, as it has been noticed in section II 3), the variance of U_n before change in real situation is far larger than that of U'_n and U''_n (see section IV 2)) ; second, as it has been outlined in sections II 1) and 2), this statistics U_n allows the detection of a smaller range of changes.

3) Performances of Hinkley's stopping-time applied to the new cusum tests when the models are known

a) The algorithm

When the two models before and after change are known, it is of common use to consider the "symmetrical" situation in which the cusum statistics has opposite drifts before and after the change. In this framework, it is obvious from (11) and (14) that the two statistics T'_n and T''_n are equivalent. Actually if we consider the symmeterized statistics :

$$\check{T}_n = T'_n + \frac{1}{2}\left[I\,(g_n^o \mid g_n^1) - I\,(g_n^1 \mid g_n^o)\right] \tag{24}$$

and :

$$\tilde{T}_n = - T''_n - \frac{1}{2}\; J\,(g_n^o\,,\,g_n^1) \qquad , \tag{25}$$

which are such that :

$$(26)\qquad \begin{cases} \mathbb{E}_{H_o}(\check{T}_n \mid Y^{n-1}) = -\frac{1}{2}\, J(g_n^o, g_n^1) & \text{(from (16))} \\ \mathbb{E}_{H_1}(\check{T}_n \mid Y^{n-1}) = +\frac{1}{2}\, J(g_n^o, g_n^1) & \text{(from (18))} \end{cases}$$

and :

$$(27)\qquad \begin{cases} \mathbb{E}_{H_o}(\tilde{T}_n \mid Y^{n-1}) = -\frac{1}{2}\, J(g_n^o, g_n^1) & \text{(from (17))} \\ \mathbb{E}_{H_1}(\tilde{T}_n \mid Y^{n-1}) = +\frac{1}{2}\, J(g_n^o, g_n^1) & \text{(from (19))} \end{cases},$$

it is also obvious from the definitions of T'_n and T''_n that :

$$T''_n = -I(g_n^o \mid g_n^1) - T'_n$$

and thus : $\check{T}_n = \tilde{T}_n$.

In other words, in this ideal situation where the two models before and after change are known, and where therefore it is always possible, without restricting the generality of the problem, to consider the "symmetrical" case, the two cumulative sum tests are identical, and proceeds as follows :

$$\tilde{U}_n = \sum_{i=1}^{n} \tilde{T}_i$$

where :

$$(28)\qquad 4\,\tilde{T}_n = \frac{\sigma_o^2}{\sigma_1^2} - \frac{\sigma_1^2}{\sigma_o^2} + \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_1^2}\right)\left[(e_n^o)^2 - (e_n^1)^2\right] + 2\left(\frac{1}{\sigma_o^2} - \frac{1}{\sigma_1^2}\right) e_n^o\, e_n^1 .$$

The purpose of this paragraph is the analysis of the behavior of Hinkley's cumulative test for detecting the change in the drift of $\tilde{U}_n$. (See Appendix II).

Let us first notice that $\tilde{U}_n$ has a symmetrical behavior (opposite drifts before and after change) but only as far as conditional expectations are concerned (see (27)). As it has been pointed out before, this is no longer true for the actual drifts of the cusum tests, which are the expectations of the previous ones with respect to the joint probability laws of the observations, before and after the change. This point will be further investigated in Section IV.

It has to be emphasized that the same kind of algorithm (cusum test with Hinkley's stopping-time) can be used in the case where the model after the change is unknown, but where the one-dimensional subspace in which the vector parameter A changes is known : I.V. NIKIFOROV [29] has shown that in this case the increment of the cusum should the corresponding linear combination of the derivatives of the conditional likelihood with respect to each of the AR parameters. But he used it only for jumps in the mean of the AR process. This direct use of Hinkley's test does not hold for jumps occuring in subspaces of dimension greater than 1.

b) The simulation study

It will give some results concerning their mean time between false alarms (F) and mean time delay for detection (R) ; in next section, the same performances will be analyzed from a theoretical point of view.

Seven models of order 3 have been chosen for this simulation, thus allowing 42 different situations of changes ; furthermore, in each case, the variances σ_o^2 and σ_1^2 of the excitation have been allowed to take two possible values : 1,10 ; there fore, each change in the AR parameters is coupled or not with a change in the variance. The seven models, described by their reflection or autoregressive coefficients, are the following ones :

Models	Reflection coefficients			Autoregressive coefficients		
	k_1	k_2	k_3	a_1	a_2	a_3
I	0.9	- 0.7	- 0.2	1.67	- 1.01	0.2
II	0.9	- 0.5	- 0.04	1.33	- 0.45	- 0.04
III	0.7	- 0.2	0.06	0.85	- 0.25	0.06
IV	- 0.9	0.5	0.8	- 0.85	0.86	0.8
V	- 0.9	0.5	0.4	- 0.65	0.68	0.4
VI	- 0.9	0.5	0.1	- 0.5	0.55	0.1
VII	- 0.9	0.3	0.05	- 0.65	0.33	0.05

The corresponding signals are shown on figure n° 12 (2000 samples) .

The cepstral distances (in dB) between these models, which have been computed with the aid of 100 coefficients, are as follows :

I						
II	0.51					
III	1.23	0.83				
IV	3.85	3.38	2.97			
V	3.42	2.94	2.46	0.72		
VI	3.17	2.71	2.17	1.13	0.44	
VII	3.35	2.89	2.89	1.20	0.56	0.30
	I	II	III	IV	V	VI

and it should be noticed that, as some of them are rather small, the detector is activated in somewhat difficult conditions.

No identifying filter is used ; the innovations e_n^o and e_n^1 are directly computed from the observations $y_n, y_{n-1}\ldots$ and the true coefficients.

Most of the mean values (except for the high thresholds) which will be mentioned in the sequel have been computed with the aid of 100 events ; i.e. 101 false alarms for computing the mean time between false alarms, and 101 alarms occuring after the true change time (which is chosen at random for each experiment) for computing the mean time delay for detection. (we recall that, as far as delay for detection is concerned, the most relevant parameter is the mean time delay for detection under the hypothesis that no false alarm occurs before the true change time. See [2] for further details).

α) <u>Changes in the AR parameters only</u>

In this case, the two test statistics $\check{U}_n$ and $\tilde{U}_n$ are equal to U'_n and U''_n, i.e. :

$$\check{U}_n = \tilde{U}_n = \frac{1}{2} \sum_{k=1}^{n} \frac{(e_k^o)^2 - (e_k^1)^2}{2\,\sigma^2} \ .$$

The two mean times are not affected by the value of $\sigma^2 = \sigma_o^2 = \sigma_1^2$ (1,10).

The chosen thresholds h were varying between 0.5 and 4. The results show that 4 is a convenient threshold, which gives a mean time between false alarms of order of magnitude 1000, except when the change is small (cepstral distance between the two models less than 1) in which case h = 3 seems to be more appropriate.

As it could be expected for, the mean time delay for detection is higher when the distance is smaller : it is equal to some units when the distance is greater than 2, and to some dizens when the distance is less than 1. Let us notice that this cepstral distance is obviously not sufficient for predicting the behavior of the detector : the position of the poles seems to be of importance, especially for the small distances (d < 1) ; this can be seen by the "dissymetry" of the results when the two models are exchanged. However, it should be emphasized that, when using the last time at which the minimum of $\tilde{U}_n$ has been reached as an estimate of the change time, this problem of dissymetry of the detection is no longer crucial : this last estimate is more accurate (bias of the estimate equal to 1 or 2 and having a small variance in most cases), and so the mean bias of the estimate of the change time is not too different when the jump occurs from signal A to signal B and in the reverse case.

β) Changes in both the AR parameters and the variance

In this case, $\check{U}_n$ and $\tilde{U}_n$ are computed with the aid of (28), i.e. :

• $$\tilde{U}_n = \sum_{i=1}^{n} \tilde{T}_i$$

with :
$$4\,\tilde{T}_n = \frac{\sigma_o^2}{\sigma_1^2} - \frac{\sigma_1^2}{\sigma_o^2} + \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_1^2}\right)\left((e_n^o)^2 - (e_n^1)^2\right) + 2\left(\frac{1}{\sigma_o^2} - \frac{1}{\sigma_1^2}\right) e_n^o\, e_n^1$$

and the threshold for Hinkley's test should be chosen at least according to $\frac{\sigma_o^2}{\sigma_1^2}$ (or $\frac{\sigma_1^2}{\sigma_o^2}$). This point will be explained in section IV from a theoretical point of view.

As in the previous case, the same remark concerning the dissymetry holds : when using the time of minimum as an estimate of the change time, one obtains a reasonable bias of the estimate (some units), and comparable accuracy when exchanging the two signals.

4) Importance of the chosen AR order

As it has been mentioned at the beginning of section II, the problem of the detection of a change in the order has not been addressed. Furthermore, in case where the order of the AR model is not known, an under-estimation of that order can give rise to a poor detection. However, it seems that the use of identifying AR filters in lattice form may prevent from a too drastic situation, if the under-estimation of the order is "reasonable" of course. In order to outline this fact, one experiment has been done (but a more extended analysis is still to be done) : a signal has been generated with the aid of an AR filter of order 5, with reflection coefficients :

$$0.7\ ,\ -\ 0.2\ ,\ 0.06\ ,\ -\ 0.8\ ,\ 0.3$$

until time 1300, and afterwards the fourth reflection coefficient has been taken to - 0.5. The cumulative sum U''_n has been computed, with the aid of identifying filters of order 5, then 4, then 3 (see Figures n° 13a, 13b, 13c respectively). It can be seen that the behavior of the detector is not too much affected by under-estimation of the order, provided that the chosen order is at least equal to the "level" at which the jump occurs (here 4). But with order 3, one could have obtained false alarms with reasonable thresholds (here, the threshold 35 is very high with respect to the behavior of U''_n before change in Figures n° 13a and b), and furthermore the behavior of U''_n after change is opposite to its normal behavior : it is first decreasing.

IV THEORETICAL STUDY

The purpose of this section is double. First, we shall use the results of [2b] to derive the performances (mean time between false alarms, mean time delay for detection) of Hinkley's stopping-time applied to $\check{U}_n$ or $\tilde{U}_n$ (see formula (24)) in the ideal situation of hypothesis testing between two known conditional laws g_n^o and g_n^1. Second, we shall analyze the asymptotic behavior of U_n, U'_n, U''_n (see formulas (8), (11)and (14)) under the hypothesis that no change occurs and for the real implementation of the algorithm, in order to explain the significant qualitative difference between their behaviors which clearly appeared for the examples shown in Section III.

1) Performances of Hinkley's stopping-time applied to $\tilde{U}_n$ in the ideal case

In order to use the theoretical results derived in [2b] concerning the mean time between false alarms and the mean time delay for detection of Hinkley's cucum test for jumps in mean (see Appendix II), we have to derive first the drifts and variances of the diffusion process equivalent to $\tilde{U}_n = \sum_{i=1}^{n} \tilde{T}_i$ (see (24), i.e. its expectations and variances under the joint probability laws of the observations before and after the jump.

It should be emphasized, however, that these computations would basically be done from a theoretical point of view and should be used with caution for the following reason : the asymptotic point of view underlying the theoretical derivation of the mean time between false alarms and the mean time delay for detection of Hinkley's test in the independent case, is much less realistic in the autoregressive case especially as far as the delay for detection is concerned, because short delays are of interest and therefore the asymptotic behavior of the cusum is far from being reached in the autoregressive case. Actually, this remark has to be related to the following fact : the Laplace transform of the pair (T_h , M_{T_h}) (where T_h is the detection time) which has been used in [2b] for deriving the mean time delay for detection and mean time between false alarms in the independent case, could not be computed in the AR gaussian case.

Therefore, these mean values will not be computed. Only the drifts before and after change will be derived.

This may further justify the simulation study which has been reported in the paragraph III 3). Some results concerning the false alarms in the real implementation will be given in the next paragraph.

From (28) :

$$4\,\tilde{T}_n = \frac{\sigma_o^2}{\sigma_1^2} - \frac{\sigma_1^2}{\sigma_o^2} + \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_1^2}\right)\left[(e_n^o)^2 - (e_n^1)^2\right] + 2\left(\frac{1}{\sigma_o^2} - \frac{1}{\sigma_1^2}\right) e_n^o\, e_n^1$$

Before the jump, i.e. under H_o, the expectation of $\tilde{T}_n$ with respect to the joint probability law of the observations is :

$$\text{(29)}\quad \mathbb{E}_{H_o}(\tilde{T}_n) = \frac{1}{4}\left[\frac{\sigma_o^2}{\sigma_1^2} - \frac{\sigma_1^2}{\sigma_o^2} + \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_1^2}\right)\left(\mathbb{E}_{H_o}(e_n^o)^2 - \mathbb{E}_{H_o}(e_n^1)^2\right.\right.$$

$$\left.+\, 2\left(\frac{1}{\sigma_o^2} - \frac{1}{\sigma_1^2}\right) \mathbb{E}_{H_o}(e_n^o\, e_n^1)\right]$$

Before the change, $\{e_n^o\}_n$ is the sequence of true innovations ; thus :

$$\text{(30)}\qquad \mathbb{E}_{H_o}(e_n^o)^2 = \sigma_o^2 \; .$$

On the other hand, if :

$$A_i(z) = \sum_{k=1}^{p} a_k^{(i)} z^k \qquad (i = 0,1) \quad ,$$

we can write :

$$e_n^1 = [1 - A_1(B)] \, y_n$$

$$= \frac{1 - A_1(B)}{1 - A_o(B)} \, e_n^o$$

(where B is the shift operator : $By_n = y_{n-1}$).

Thus :

$$\mathbb{E}_{H_o}(e_n^1)^2 = \frac{\sigma_o^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1 - A_1(e^{j\theta})}{1 - A_o(e^{j\theta})} \right|^2 d\theta \ .$$

Considering the Taylor's expansion :

(31) $$\frac{1 - A_1(z)}{1 - A_o(z)} = 1 + \sum_{k=1}^{+\infty} c_k^{1/0} z^k \quad ,$$

it comes :

(32) $$\mathbb{E}_{H_o}(e_n^1)^2 = \sigma_o^2 \left(1 + \sum_{k=1}^{+\infty} (c_k^{1/0})^2\right) \ .$$

Furthermore :

(33) $$\mathbb{E}_{H_o}(e_n^o \, e_n^1) = \frac{\sigma_o^2}{2\pi} \int_{-\pi}^{\pi} \frac{1 - A_1(e^{j\theta})}{1 - A_o(e^{j\theta})} \, d\theta = \sigma_o^2$$

Finally, from (29) - (32), the drift before change is :

(34) $$\mathbb{E}_{H_o}(\tilde{T}_n) = \frac{1}{4} \left[2 - \left(\frac{\sigma_o^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_o^2}\right) - \left(1 + \frac{\sigma_o^2}{\sigma_1^2}\right) \sum_{k=1}^{+\infty} (c_k^{1/0})^2 \right] < 0.$$

After the jump, i.e. under H_1, the expectation is similarly :

$$(35)\qquad \mathbb{E}_{H_1}(\tilde{T}_n) = \frac{1}{4}\left[-2 + \left(\frac{\sigma_o^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_o^2}\right) + \left(1 + \frac{\sigma_1^2}{\sigma_o^2}\right)\sum_{k=1}^{+\infty} (c_k^{0/1})^2\right] > o ,$$

where the coefficients $c_k^{0/1}$ are given by :

$$(36)\qquad \frac{1 - A_o(z)}{1 - A_1(z)} = 1 + \sum_{k=1}^{+\infty} c_k^{0/1} z^k .$$

It can be seen, from these two expectations, that the "dissymetry" of the cusum test $\tilde{U}_n$ (see section II 2)) i.e. the fact that $\mathbb{E}_{H_o}(\tilde{T}_n) \neq - \mathbb{E}_{H_1}(\tilde{T}_n)$, depends only upon the last term :

$$(37)\qquad \left(1 + \frac{\sigma_o^2}{\sigma_1^2}\right) \sum_{k=1}^{+\infty} (c_k^{1/0})^2 - \left(1 + \frac{\sigma_1^2}{\sigma_o^2}\right) \sum_{k=1}^{+\infty} (c_k^{0/1})^2 ;$$

in the case where $\sigma_o = \sigma_1$, the summations begin at time n = 2 because $c_n^{1/0} = - c_n^{0/1}$. This expression explains the results obtained by simulation in section II 3).

It should be noticed that the same series explain the slight difference in the drifts of the cusum test in real implementation when the simulated signals before and after the jump are exchanged.

2) Asymptotic behavior, for the real implementation, of U_n , U'_n , U''_n when no change occurs.

Recall that we consider here the real situation where the two identified models M'_1 and M_2 are identical when no change occurs. We shall work under the assumption that, on one hand, the first filter (sequential one) has converged, and thus that e_n^o is the true innovation and (a_i^o $(1 \leq i \leq p)$, σ_o^2) are the true parameters ; and, on the other hand, that the parameters ($\hat{a}_i^1$ $(1 \leq i \leq p)$, $\hat{\sigma}_1^2$) estimated by the second filter (sliding block) are slowly varying.

It has been seen (3) that :

$$z_n = \frac{1}{\sqrt{2n}} \sum_{k=1}^{n} \left(\frac{e_k^2}{\sigma_o^2} - 1\right)$$

is asymptotically distributed as a gaussian law $\mathcal{N}(0,1)$. Using (7) :

$$U_n = \sqrt{\frac{n}{2}}\; z_n$$

and thus :

$$E(U_n) = o$$

(48) $$\boxed{\mathrm{var}\,(U_n) = \frac{n}{2}} \; .$$

On the other hand, let : $\Delta e_{n,N} = e_n^1 - e_n^o$ be the "innovation error" for the sliding block filter, and :
$\Delta\sigma^2_{n,N} = \hat{\sigma}_1^2 - \sigma_o^2$ the estimated variance error for the same filter, which are small by assumption and depend upon both the current time n and the block size N.

$$\text{Then : } T'_n = \frac{1}{2} \operatorname{Log} \frac{\sigma_o^2}{\hat{\sigma}_1^2} + \frac{(e_n^o)^2}{2\,\sigma_o^2} - \frac{(e_n^1)^2}{2\,\hat{\sigma}_1^2}$$

$$= -\frac{1}{2} \operatorname{Log}\left(1 + \frac{\Delta\sigma^2_{n,N}}{\sigma_o^2}\right) + \frac{(e_n^o)^2 - (e_n^1)^2}{2\,\sigma_o^2}$$

$$+ \frac{(e_n^1)^2}{2} \left[\frac{1}{\sigma_o^2} - \frac{1}{\sigma_o^2 + \Delta\sigma^2_{n,N}}\right]$$

(49) $$\simeq \frac{\Delta\sigma^2_{n,N}}{2\,\sigma_o^2} \left[\frac{(e_n^o)^2}{\sigma_o^2} - 1\right] - \frac{e_n^o\,\Delta e_{n,N}}{\sigma_o^2} \; ,$$

by neglecting the "second" order terms.

As $\Delta\sigma^2_{n,N}$ is slowly varying with respect to e^o_n, the heuristic argument which can be found in Appendix III shows that $\Delta\sigma^2_{n,N}$ is approximately independent of e^o_n, and thus :

$$E(\Delta\sigma^2_{n,N}\ (\frac{(e^o_n)^2}{\sigma^2_o} - 1)) = E(\Delta\sigma^2_{n,N}) .\ E(\frac{(e^o_n)^2}{\sigma^2_o} - 1)$$
$$= o \ ,$$

because of the second term of the product.

On the other hand, as e^o_n is assumed to be the exact innovation, the error $\Delta e_{n,N}$ which is in the space spanned by Y^{n-1}, is independant of e^o_n. Therefore :

$$E(e^o_n\ \ \Delta e_{n,N}) = E(e^o_n)\ \ E(\Delta e_{n,N})$$
$$= o \ ,$$

and $E(T'_n) = o$. Thus : $E(U'_n) = \sum_{i=1}^{n} E(T'_i) = o$.

Furthermore, $\mathrm{var}(U'_n) = E\left(\sum_{k=1}^{n} \left[\frac{\Delta\sigma^2_{k,N}}{2\ \sigma^2_o}\ (\frac{(e^o_k)^2}{\sigma^2_o} - 1) - \frac{e^o_k\ \Delta e_{k,N}}{\sigma^2_o}\right]\right)^2$

(50) $$\leq 2\left(E \sum_{k=1}^{n} \frac{\Delta\sigma^2_{n,N}}{2\ \sigma^2_o}\ (\frac{(e^o_k)^2}{\sigma^2_o} - 1)\right)^2 + 2\left(E \sum_{k=1}^{n} \frac{e^o_k\ \Delta e_{k,N}}{\sigma^2_o}\right)^2$$

by Holder's inequality.

Let us consider the second term of this sum. Using the independence pro perty of the sequence of true innovations (e^o_n), we have for $k > j$:

$$E(e_k^o \, \Delta e_{k,N} \, e_j^o \, \Delta e_{j,N}) = E(e_k^o) \quad E(\Delta e_{k,N} \, e_j^o \, \Delta e_{j,N})$$

$$= o$$

because of the first term of the product. Thus :

$$E\left(\sum_{k=1}^{n} e_k^o \, \Delta e_{k,N}\right)^2 = \sum_{k=1}^{n} E(e_k^o \, \Delta e_{k,N})^2$$

$$= \sum_{k=1}^{n} \mathrm{var}(e_k^o) \quad \mathrm{var}(\Delta e_{k,N})$$

because $\Delta e_{k,N}$ is independant of e_k^o.

But : $\Delta e_{k,N} = e_k^1 - e_k^o$

$$= (\hat{A}^1 - A^o) \; Y^{k-1} \; .$$

As the estimation error $\Delta\hat{A}_{k,N} = \hat{A}^1 - A^o$ is slowly varying with respect to the observations Y^{k-1}, we can again consider (Appendix III) that $\Delta\hat{A}_{k,N}$ is independent of Y^{k-1}.

$$E(\Delta e_{k,N})^2 = E(\Delta\hat{A}_{k,N} \; Y^{k-1})^2$$

$$= E(E[(\Delta\hat{A}_{k,N} \; Y^{k-1})^2 \mid \Delta\hat{A}_{k,N}])$$

$$= E(\Delta\hat{A}_{k,N} \; E(Y^{k-1} \; Y^{k-1'} \mid \Delta\hat{A}_{k,N}) \; \Delta\hat{A}'_{k,N})$$

$$= E(\Delta\hat{A}_{k,N} \; E(Y^{k-1} \; Y^{k-1'}) \; \Delta\hat{A}'_{k,N})$$

by independence of $\Delta\hat{A}_{k,N}$ and Y^{k-1}.

Thus : $E(\Delta e_{k,N})^2 = E(\Delta\hat{A}_{k,N} \; R_o \; \Delta\hat{A}'_{k,N})$,

where R_o is the covariance matrix of the observations (y_n) before change.

It is known that the estimation error $\Delta\hat{A}_{k,N} = \hat{A}^1 - A^o$, based upon the N observations of the block, is asymptotically distributed as a gaussian law $\mathbf{N}(o, \frac{\sigma_o^2 R_o^{-1}}{N})$; therefore : $\mathrm{var}(\Delta e_{k,N}) = \frac{\sigma_o^2}{N}$, and :

$$E\left(\sum_{k=1}^{n} e_k^o \Delta e_{k,N}\right)^2 = \sum_{k=1}^{n} \sigma_o^2 \frac{\sigma_o^2}{N} = \frac{n}{N}\sigma_o^4 . \tag{51}$$

Consider now the first term of the sum in (50). Using the fact that e_k^o is independent of $\Delta\sigma^2_{k,N}$ because they are varying with different speeds (see Appendix III), and that, for k > j, e_k^o is independent of e_j^o and $\Delta\sigma^2_{j,N}$ because of the independence property of the sequence of true innovations (e_n^o), we have, for k > j :

$$E\left(\Delta\sigma^2_{k,N}\left(\frac{(e_k^o)^2}{\sigma_o^2} - 1\right) \Delta\sigma^2_{j,N}\left(\frac{(e_j^o)^2}{\sigma_o^2} - 1\right)\right)$$

$$= E\left(\frac{(e_k^o)^2}{\sigma_o^2} - 1\right) \cdot E\left(\Delta\sigma^2_{k,N}\, \Delta\sigma^2_{j,N}\left(\frac{(e_j^o)^2}{\sigma_o^2} - 1\right)\right)$$

$$= o$$

because of the first term of the product. Thus :

$$E\left(\sum_{k=1}^{n} \frac{\Delta\sigma^2_{k,N}}{2\sigma_o^2}\left(\frac{(e_k^o)^2}{\sigma_o^2} - 1\right)\right)^2 = \sum_{k=1}^{n} E\left(\frac{\Delta\sigma^2_{k,N}}{2\sigma_o^2}\left(\frac{(e_k^o)^2}{\sigma_o^2} - 1\right)\right)^2$$

$$= \frac{1}{4\sigma_o^4}\sum_{k=1}^{n} \mathrm{var}(\Delta\sigma^2_{k,N}) \cdot \mathrm{var}\left(\frac{(e_k^o)^2}{\sigma_o^2} - 1\right) \tag{52}$$

because $\Delta\sigma^2_{k,N}$ is independent of e_k^o (Appendix III).

But :

$$\mathrm{var}\left(\frac{(e_k^o)^2}{\sigma_k^2} - 1\right) = E\left(\frac{(e_k^o)^4}{\sigma_o^4} - 2\frac{(e_k^o)^2}{\sigma_o^2} + 1\right) = 2 \tag{53}$$

under gaussian assumption, and :

since : $\Delta\sigma^2_{k,N} = \hat{\sigma}^2_1 - \sigma^2_o$

$\simeq 2\,\sigma_o\;\Delta\sigma_{k,N}$,

we have :

(54) $$\mathrm{var}(\Delta\sigma^2_{k,N}) = 4\,\sigma^2_o \cdot \mathrm{var}(\Delta\sigma_{k,N}) .$$

In the AR gaussian case, it is known that :

$$\mathrm{var}(\Delta\sigma_{k,N}) = \frac{\sigma^2_o}{2N} .$$

So, by (54) :

$$\mathrm{var}(\Delta\sigma^2_{k,N}) = \frac{2\,\sigma^4_o}{N} .$$

Therefore :

(55) $$E\left(\sum_{k=1}^{n} \frac{\Delta\sigma^2_{k,N}}{2\,\sigma^2_o}\left(\frac{(e^o_k)^2}{\sigma^2_o} - 1\right)\right)^2 = \frac{1}{4\,\sigma^4_o}\sum_{k=1}^{n}\frac{2\,\sigma^4_o}{N}\cdot 2 = \frac{n}{N} ,$$

and :

(56) $\boxed{\mathrm{var}\,(U'_n) \leq \frac{4n}{N}}$, using (50), (51) and (55).

In the same manner,

$$T''_n = \frac{\sigma^2_o}{2\,\sigma^2_1}\left[2\,\frac{e^o_n\,e^1_n}{\sigma^2_o} - \left(1 + \frac{\sigma^2_1}{\sigma^2_o}\right)\frac{(e^o_n)^2}{\sigma^2_o} + \left(\frac{\sigma^2_1}{\sigma^2_o} - 1\right)\right]$$

$$= \frac{1}{2\left(1 + \frac{\Delta\sigma^2}{\sigma^2_o}\right)}\left[2\,\frac{e^o_n\,(e^o_n + \Delta e_n)}{\sigma^2_o} - \left(2 + \frac{\Delta\sigma^2}{\sigma^2_o}\right)\frac{(e^o_n)^2}{\sigma^2_o} + \frac{\Delta\sigma^2}{\sigma^2_o}\right]$$

(57) $$T''_n \simeq \frac{e^o_n\,\Delta e_{n,N}}{\sigma^2_o} - \frac{\Delta\sigma^2_{n,N}}{2\,\sigma^2_o}\left(\frac{(e^o_n)^2}{\sigma^2_o} - 1\right) .$$

In other words, $T''_n \simeq - T'_n$ (compare (49) and (57)), and therefore U''_n has approximately the same variance as U'_n, which is $\frac{1}{N}$ the order of magnitude of the variance of U_n. This explains the much better behavior, before change, of our two test statistics U'_n and U''_n with respect to the more classical one U_n.

Since U'_n and $- U''_n$ have the same behavior before change (for <u>real</u> implementation) and that, by (18) and (19), $- U''_n$ has <u>a larger drift</u> than U'_n after change, <u>the cusum test U''_n will be preferred to U'_n</u> .

Furthermore, the properties of Hinkley's stopping-time when applied to this statistics $- U''_n$ (in real implementation) can be derived from the previous computations. Indeed, with an a priori fixed drift $\delta < o$ before jump, the mean time between false alarms is given by :

$$F = \frac{1}{\delta} \left[\frac{1}{2\gamma} (e^{2\gamma h} - 1) - h\right]$$

where $\gamma = \dfrac{\delta}{var(T'')}$. See Appendix II.

The previous computations have shown that var(T") has $\frac{4}{N}$ as order of magnitude. Therefore :

$$\boxed{F = \frac{1}{\delta} \left[\frac{2}{\delta N} (e^{\frac{\delta N h}{2}} - 1) - h\right]}$$

where h is the threshold and N the size of the window used for identifying M_1.

For the same reason as in the previous paragraph, the mean time delay for detection will not be computed. Let us notice that the drift of $- U''_n$ after change is :

$$- \delta + 2\, \mathbb{E}_{H_1} (\tilde{T}_n)$$

because of (19) and (26) ; i.e. from (35) :

$$- \delta + \frac{1}{2} [- 2 + (\frac{\sigma_o^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_o^2}) + (1 + \frac{\sigma_1^2}{\sigma_o^2}) \sum_{k=1}^{\infty} (c_k^{0/1})^2] .$$

V. CONCLUSION

The problem of sequential segmentation of nonstationary digital signals, using spectral analysis, has been addressed. In this framework, the sequential detection of abrupt changes in spectral characteristics is of interest. The limitations of some classical approaches, based upon the innovations of one AR model, have been outlined, and some new algorithms have been derived. The underlying idea of these algorithms is the use of two AR models : the first one is a global (or long term)estimated model which is updated recursively, the second one is a local (or short term) one updated via a sliding block. Changes are detected when a suitable distance beween these two models is too high. Three "distance" measures are considered ; cepstral distance, log-likelihood ratio and Kullback's divergence between conditional probability laws. Hinkley's stopping-time may be coupled to these last two tests in order to give a better estimate of the change-time ; the properties of these detectors have been theoretically analyzed. Moreover, the three algorithms have been compared both via a simulation study and from a theoretical point of view, and the improvement with respect to the classical approach is demonstrated.

APPENDIX 1

Sequential approximated least squares filter in lattice form.

The purpose of this appendix is the presentation of the equations of the sequential filter we use for identifying model M'_o (see Figure n° 1).

Let the model be :

$$(1) \qquad y_t = \sum_{i=1}^{p} a_i \, y_{t-i} + e_t(p+1) \quad ,$$

and $(e_t(n) \; ; \; 1 \leq n \leq p+1)$ and $(f_t(n) \; ; \; 1 \leq n \leq p+1)$ be the forward and backward innovations respectively. In the scalar case, these two innovations have the same variances :

$$^{(e)}\sigma_t^2(n) = {}^{(f)}\sigma_t^2(n) = \sigma_t^2(n)$$

At the first step : $(t = 1)$:

$$(2) \left\{ \begin{array}{l} e_1(1) = f_1(1) = y_1 \\ \sigma_1^2(1) = \gamma_1 \, e_1^2(1) \quad . \end{array} \right.$$

Afterwards $(t \geq 2)$ the algorithm proceeds as follows :

$$(3) \quad \begin{array}{l} \cdot \; \sigma_t^2(1) = \sigma_{t-1}^2(1) + \gamma_t \; (y_t^2 - \sigma_{t-1}^2(1)) \\ \cdot \; e_t(1) = y_t \\ \cdot \; N_t = \min\,(p,\, t-1) \end{array}$$

- Memorize the backward innovations $(f_{t-1}(n) \; ; \; 1 \leq n \leq N_{t-1} + 1)$

- For $n = 1, N_t$:

. Correlation between the forward and backward innovations :

(5) $$cor_t(n) = cor_{t-1}(n) + \gamma_{t-n}\ (e_t(n)\ f_{t-1}(n) - cor_{t-1}(n))$$

. Reflection coefficient :

(6) $$k_t(n) = \frac{2\ cor_t(n)}{{}^{(e)}\sigma_t^2(n) + {}^{(f)}\sigma_t^2(n)}$$

. Forward and backward innovations :

(7) $$e_t\ (n+1) = e_t\ (n) - k_t\ (n)\ f_{t-1}\ (n)$$
$$f_t\ (n+1) = f_{t-1}(n) - k_t\ (n)\ e_t\ (n)$$

. Variances of these innovations :

(8) $${}^{(e)}\sigma_t^2\ (n+1) = {}^{(e)}\sigma_{t-1}^2(n+1) + \gamma_{t-n}\ (e_t^2(n+1) - {}^{(e)}\sigma_{t-1}^2(n+1))$$
$${}^{(f)}\sigma_t^2\ (n+1) = {}^{(f)}\sigma_{t-1}^2(n+1) + \gamma_{t-n}\ (e_t^2(n+1) - {}^{(f)}\sigma_{t-1}^2(n+1))$$

(9) - Finally : $f_t(1) = y_t$.

In all these equations, the gains γ_t are as follows :

(10) $$\gamma_t = \gamma_o + \frac{1}{t}$$

where γ_o is a priori fixed.

For further details, see [5].

Cepstral coefficients and distance

Under the assumption that the model (1) is minimum phase, its cepstral coefficients are, by definition, the Fouriers's coefficients of the corresponding log spectrum, i.e; :

$$\mathrm{Log} \frac{\sigma^2}{\left|1 - \sum_{k=1}^{p} a_k e^{kj\omega}\right|^2} = \sum_{k=-\infty}^{+\infty} c_k e^{-kj\omega} \quad ,$$

with : (11) $c_o = \mathrm{Log}(\sigma^2)$ and $c_{-k} = c_k$, and can be recursively computed from the filter coefficients via the formulas :

$$(12) \qquad - n \; c_n - n \; a_n = \sum_{k=1}^{n-1} (n - k) \; c_{n-k} \; a_k \; (n > o)$$

From Parseval's relation, it can be seen that the mean square distance between the logarithms of two such spectra, is nothing but :

$$d = \left([c_o^{(o)} - c_o^{(1)}]^2 + 2 \sum_{k=1}^{+\infty} [c_k^{(o)} - c_k^{(1)}]^2 \right)^{\frac{1}{2}} .$$

Furthermore, a reasonable approximation to d can be obtained by truncating this last serie at least at the order of the filter.

For further details, see [10].

APPENDIX II

Hinkley's cumulative sum test

1) The algorithm

This test is devoted to the sequential detection of changes in mean for a sequence of independent observations governed by a gaussian law $\mathcal{N}(\mu_o , \sigma^2)$ before the change and $\mathcal{N}(\mu_1 , \sigma^2)$ after the change. It has been shown in [2b] that this test is optimal, with respect to the criterion : " minimize the mean time delay for detection for a given mean time between false alarms".Furthermore, it is also more robust than other tests with respect to a poor estimation of μ_o and σ^2. These two superiorities of Hinkley's test are more obvious when σ^2 is high.

In the (real) situation where μ_o and $\mu_1$are unknown, μ_o is estimated (and, possibly, σ^2) and a minimum magnitude of change ν is fixed a priori. Two tests are then activated in parallel : on one hand :

$$\left\{ \begin{array}{l} s_n = \sum_{i=1}^{n} (Y_i - \mu_o - \frac{\nu}{2}) \quad (n \geq 1) \; / \; (s_o = o) \\ \\ m_n = \min_{o \leq k \leq n} s_k \\ \\ \text{Alarm when : } s_n - m_n \geq h \end{array} \right.$$

is looking for upwards jumps ($\mu_o \rightarrow \mu_o + \nu$) ; on the other hand :

$$\left\{ \begin{array}{l} S_n = \sum_{i=1}^{n} (Y_i - \mu_o + \frac{\nu}{2}) \quad (n \geq 1) \quad (S_o = o) \\ \\ M_n = \max_{o \leq k \leq n} S_k \\ \\ \text{Alarm when : } M_n - S_n \geq h \end{array} \right.$$

is looking for downwards jumps ($\mu_o \to \mu_o - \nu$).

There are two possible estimates for the jump time: either the alarm time (possible reduced by its bias), or the instant at which the maximum of S_k (resp. the minimum of s_k) has been reached.

In our framework of detection of abrupt changes in spectral characteristics, in real implementation, the mean μ_o before jump is known to be zero for each of the cusums U_n, U'_n, U''_n (see section II 1) and 2)). Furthermore, the minimum magnitude of change ν is here related to the minimum conditional divergence between the two AR models : see formula (19) in section II 2). Some examples are shown in section III 2). In the ideal case where the models are known, the tests is simply : $\tilde{U}_n - m_n \geq h$, where $\tilde{U}_n$ is given by (28). See section III 3).

2) Theoretical results

It has been shown in [2b] that, for a change from $\mathbf{N}(\mu_o , \sigma^2)$ to $\mathbf{N}(\mu_1 , \sigma^2)$, the mean time between false alarms is :

$$F = \begin{cases} \dfrac{h^2}{\sigma^2} & \text{if} \quad \mu_o = o \\[2ex] \dfrac{1}{\mu_o}\left[\dfrac{1}{2\gamma_o}\left(e^{2\gamma_o h} - 1\right) - h\right] & \text{if} \quad \mu_o \neq o \end{cases}$$

where $\gamma_o = \dfrac{\mu_o}{\sigma^2}$,

and the mean time delay for detection, conditionally to the fact that no false alarm has occured before the jump is given by :

$$-\mu_1 R = e^{-2\gamma_0 h}\left(h - \frac{e^{\gamma_1 h}\,\mathrm{sh}(\gamma_1 h)}{\gamma_1}\right) + \frac{e^{-\gamma_1 h}}{2(\gamma_0-\gamma_1)\,\mathrm{sh}(\gamma_1 h)}\left(1 - \gamma_1 h \coth(\gamma_1 h)\right)$$

$$- \frac{e^{-2\gamma_0 h}}{2(\gamma_0-\gamma_1)\,\mathrm{sh}(\gamma_1 h)}\left(\mathrm{ch}(\gamma_1 h) - \frac{\gamma_1 h}{\mathrm{sh}(\gamma_1 h)}\right)$$

$$+ \frac{2\gamma_0}{\theta_1+2\gamma_0}\left(1-e^{-(\theta_1+2\gamma_0)h}\right)\left[\frac{1}{\gamma_1} - h\coth(\gamma_1 h) - 2\,\frac{\theta_1+\gamma_0}{\gamma_1(\theta_1+2\gamma_0)}\,e^{\gamma_1 h}\left(\mathrm{ch}(\gamma_1 h) - \frac{\gamma_1 h}{\mathrm{sh}(\gamma_1 h)}\right)\right]$$

$$+ \frac{2\gamma_0 h}{\theta_1+2\gamma_0}\,\frac{1}{\mathrm{sh}(\gamma_1 h)}\left(\mathrm{ch}(\gamma_1 h) - \frac{\gamma_1 h}{\mathrm{sh}(\gamma_1 h)}\right) e^{-(\theta_1+2\gamma_0)h}$$

$$+ \frac{\gamma_1}{4\gamma_0(\gamma_0-\gamma_1)^2}\left(\theta_1 - e^{-2\gamma_0 h}(\theta_1+2\gamma_0)\right)$$

where (19) $\gamma_0 = \dfrac{\mu_0}{\sigma^2}$, $\gamma_1 = \dfrac{\mu_1}{\sigma^2}$, $\theta_1 = \gamma_1 \dfrac{e^{-\gamma_1 h}}{\mathrm{sn}(\gamma_1 h)}$.

APPENDIX III

Instataneous independence of two jointly stationary processes which are varying at different speeds

Let us consider a process (X_n) the trajectory of which is quickly varying and a process (Z_n) the trajectory of which is slowly varying and let us assume that (X_n, Z_n) is stationary. We shall heuristically justify that the random variable X_n is independent of the random variable Z_n.

By the law of large numbers :

$$E(f(Z.)\ g(X.)) \simeq \frac{1}{KN} \sum_{i=1}^{KN} f(Z_i)\ g(X_i) \quad ,$$

where f and g are two bounded continuous fonctions. Dividing the time interval (1 ,..., KN) into K intervals of length N where Z is nearly constant, we obtain :

$$\frac{1}{KN} \sum_{i=1}^{KN} f(Z_i)\ g(X_i) \simeq \frac{1}{K} \sum_{k=1}^{K} \left[f(Z_{kN}) \left(\frac{1}{N} \sum_{i=1}^{N} g(X_{kN+i}) \right) \right]$$

$$\simeq \frac{1}{K} \sum_{k=1}^{K} [\, f(Z_{kN}) \quad .E(g(X.))]$$

$$\simeq E(f(Z.) \quad .E(g(X.))$$

using twice the law of large numbers, and therefore :

$$E\ (f(Z.)\ g(X.)) = E\ (f(Z.))\ .E\ (g(X.)) \quad .$$

ACKNOWLEDGEMENT

The authors would like to thank Pr. B. DICKINSON (Princeton Univ. N.J.) for a fruitful discussion concerning the conditional entropy approach of the classical test, which enlightened its weaknesses and led to the derivation of a new algorithm.

FIGURE CAPTION

Figure n° 1a :	One general approach for the detection of model changes.
n° 1b :	Two approaches for detecting model changes, using two models.
n° 2 :	Comparison between U_n and U''_n for an AR (1) signal : $a_1^o = 0.6$; $a_1^1 = 0.1$; $\sigma_o = \sigma_1 = 1$.
n° 3 :	Comparison between U_n and U''_n for an AR (2) signal, in the case of change of energy. $a_1^o = a_1^1 = 0.3$; $a_2^o = a_2^1 = 0.5$; $\sigma_o^2 = 4$; $\sigma_1^2 = 1/4$.
n° 4 :	Comparison between U_n and U''_n for an AR (4) signal. Changes in energy and in AR parameters. $a_1^o = 0.3$; $a_2^o = 0.5$; $a_3^o = a_4^o = o$; $a_1^1 = 0.5$; $a_2^1 = -0.3$; $a_3^1 = 0.6$; $a_4^1 = -0.5$; $\sigma_o^2 = 0.09$; $\sigma_1^2 = 0.16$.
n° 5 :	Comparison between U_n, U''_n, U'_n for an AR (3) signal (III → II) .
n° 6 :	Comparison between U'_n and U''_n for a "small" change in an AR (3) model (I → II).
n° 7 :	Behavior of U'_n for a change in an AR (3) model coupled with a decrease in energy (III → I ; $\sigma_o^2 = 10$; $\sigma_1^2 = 1$) .
n° 8 :	Behavior of U''_n when the covariance method is used in the sliding block. (III → II)
n° 9 :	Behavior of the cepstral distance when autocorrelation (n° 9a) or covariance (n° 9b) method is used in the sliding block. (I → III).

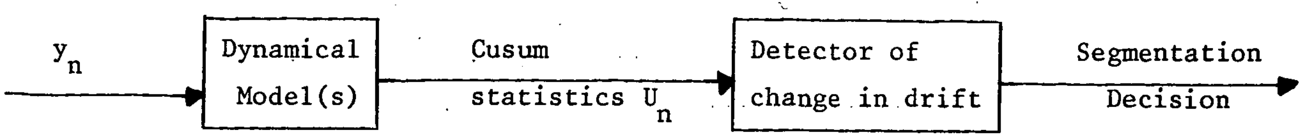


Figure n° 1a : One general approach for the detection of model changes.

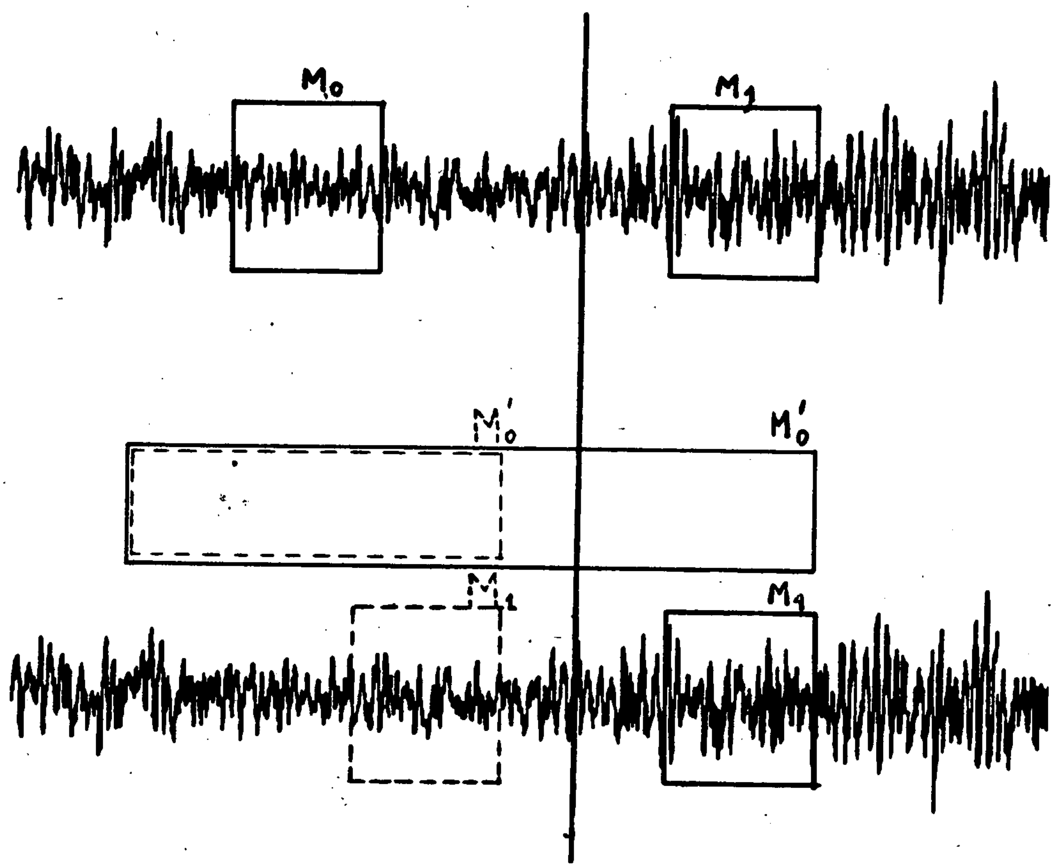


Figure n° 1b : Two approaches for sequential detection of model changes, using two models.

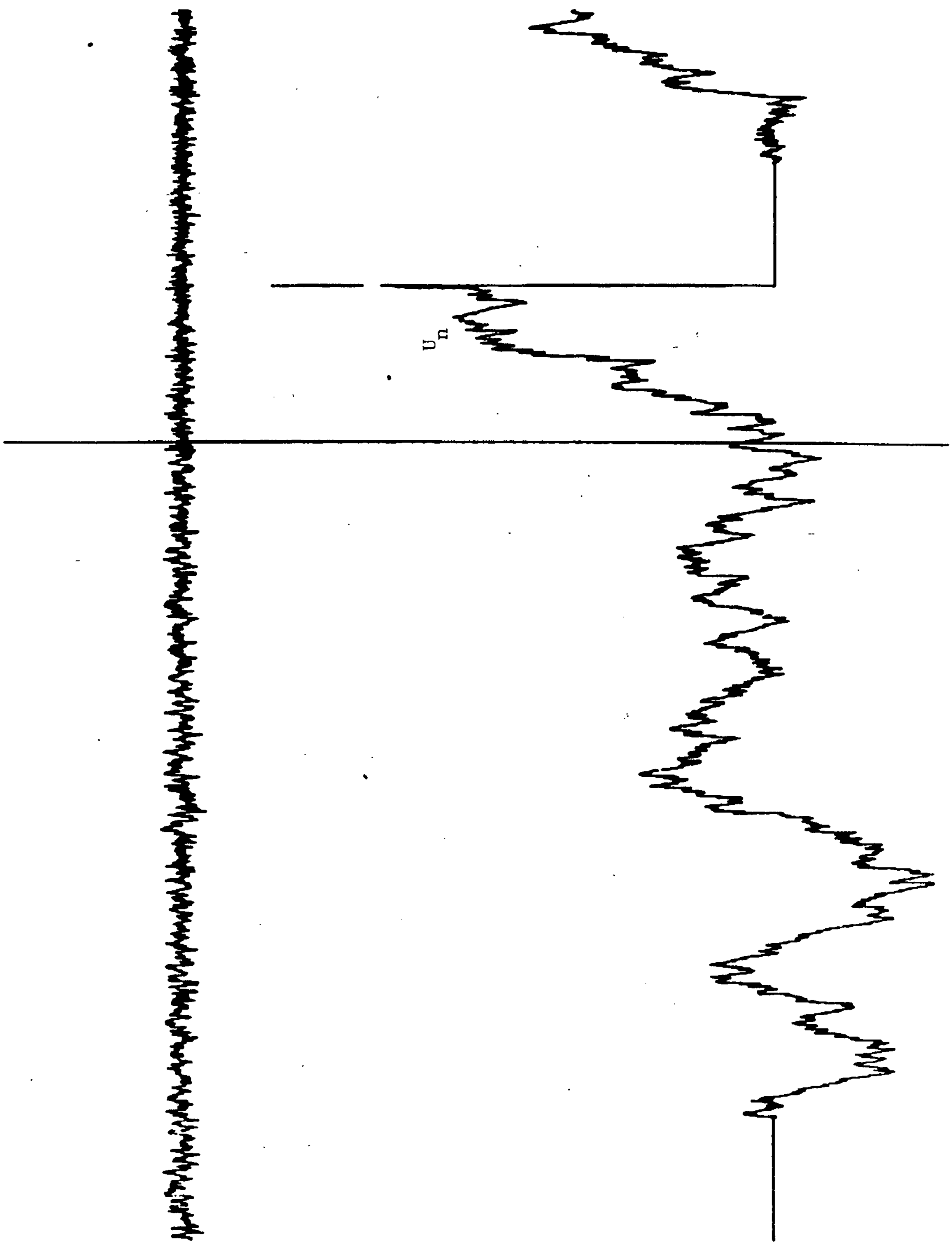


Figure n° 2a

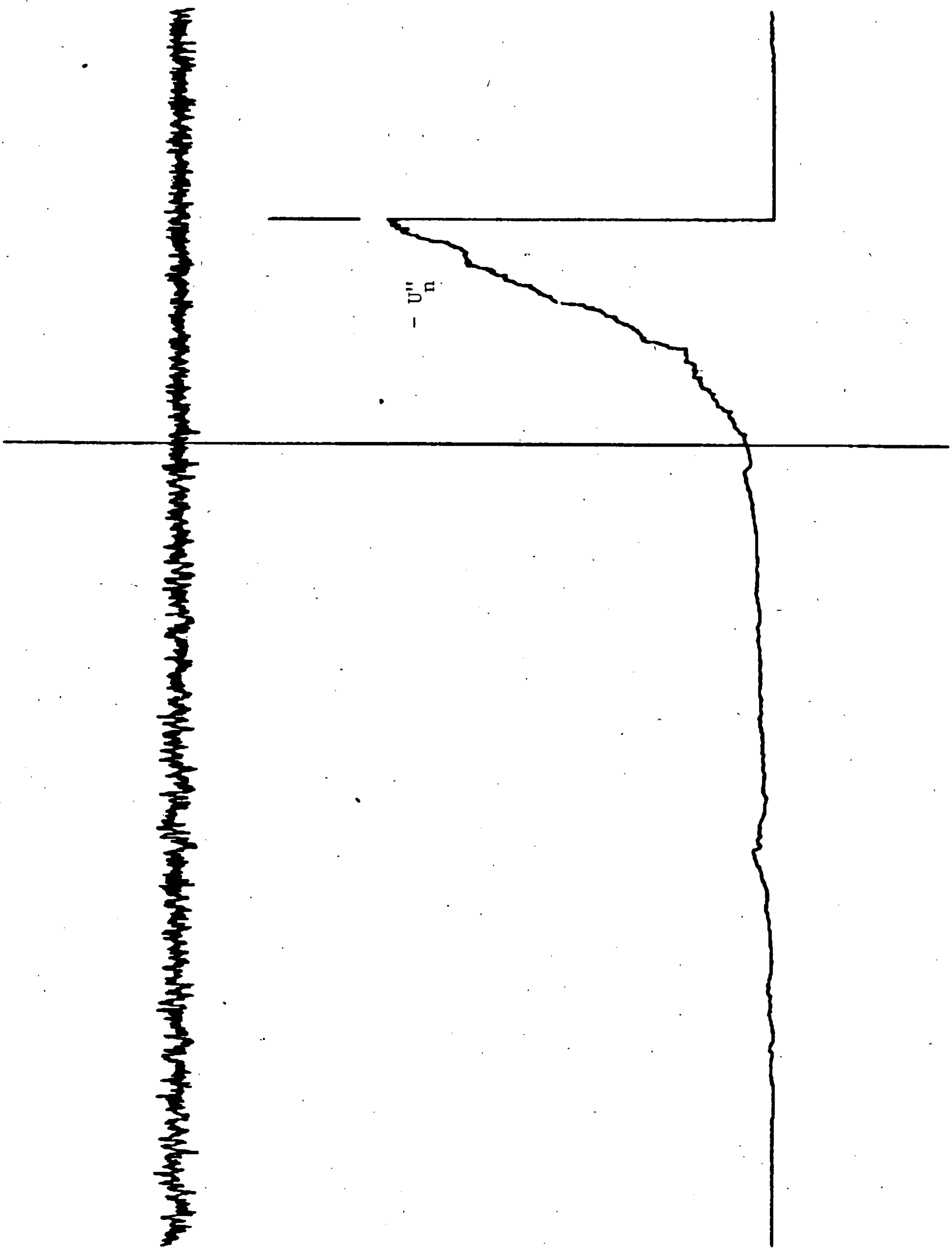


Figure n° 2b

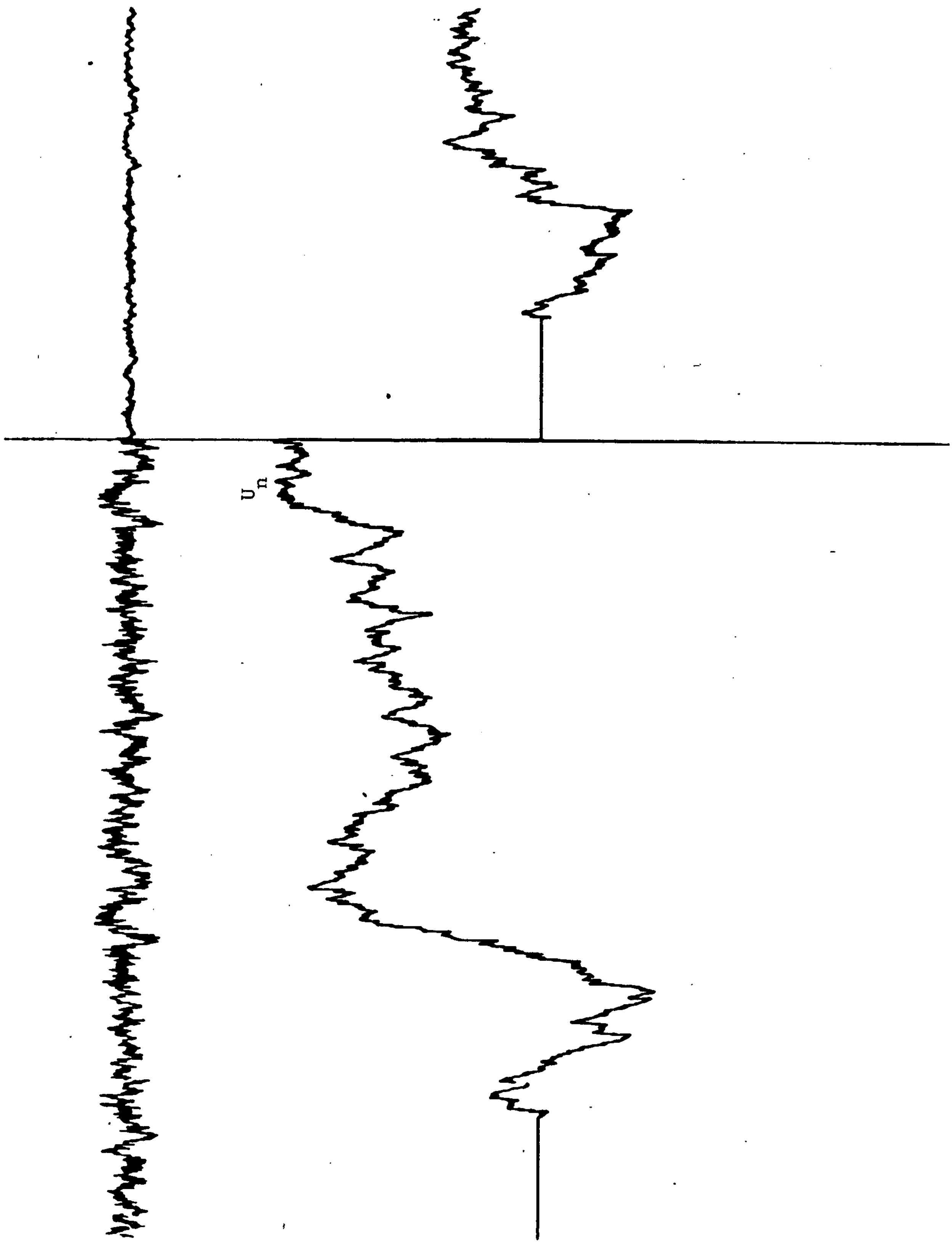


Figure n° 3a

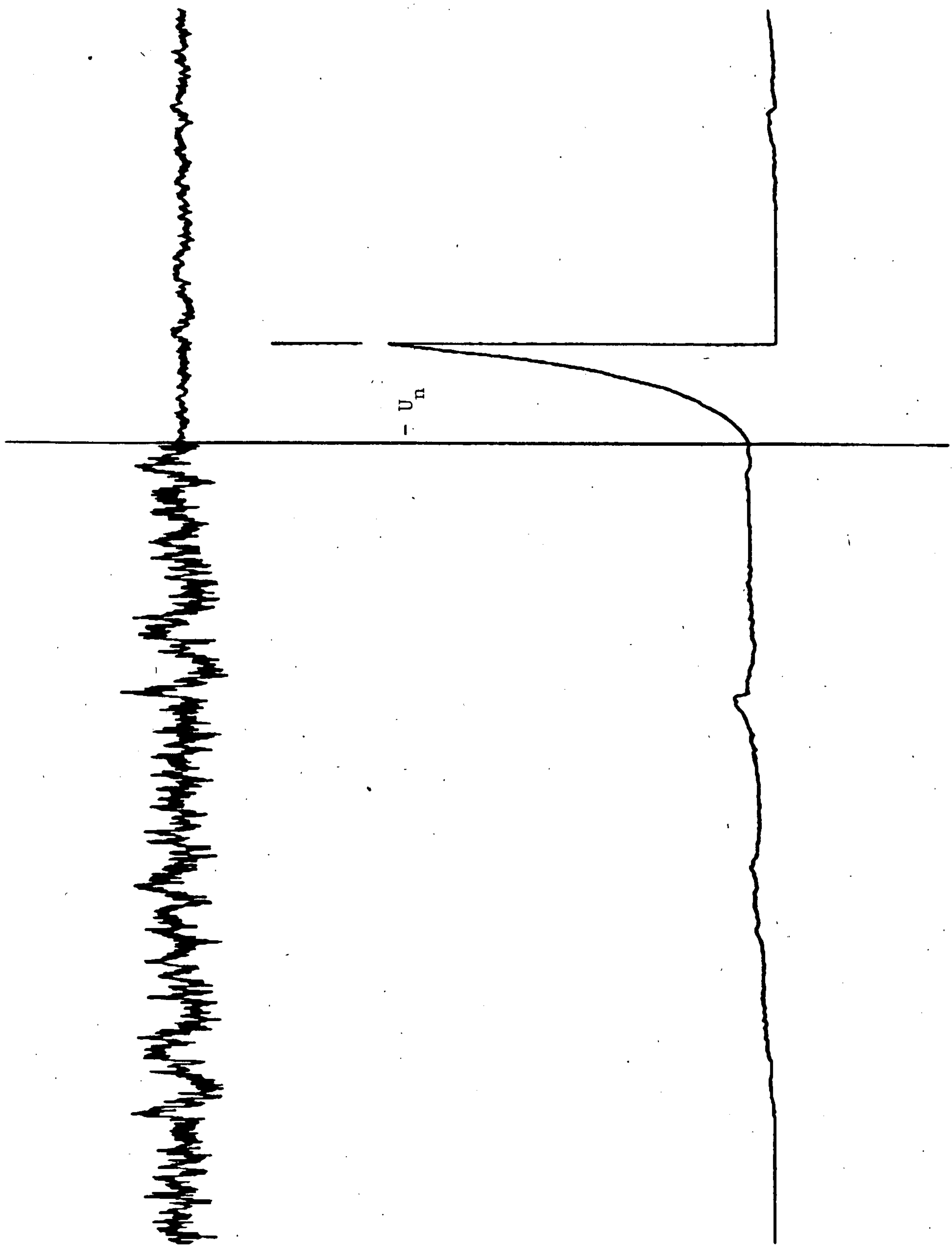


Figure n° 3b

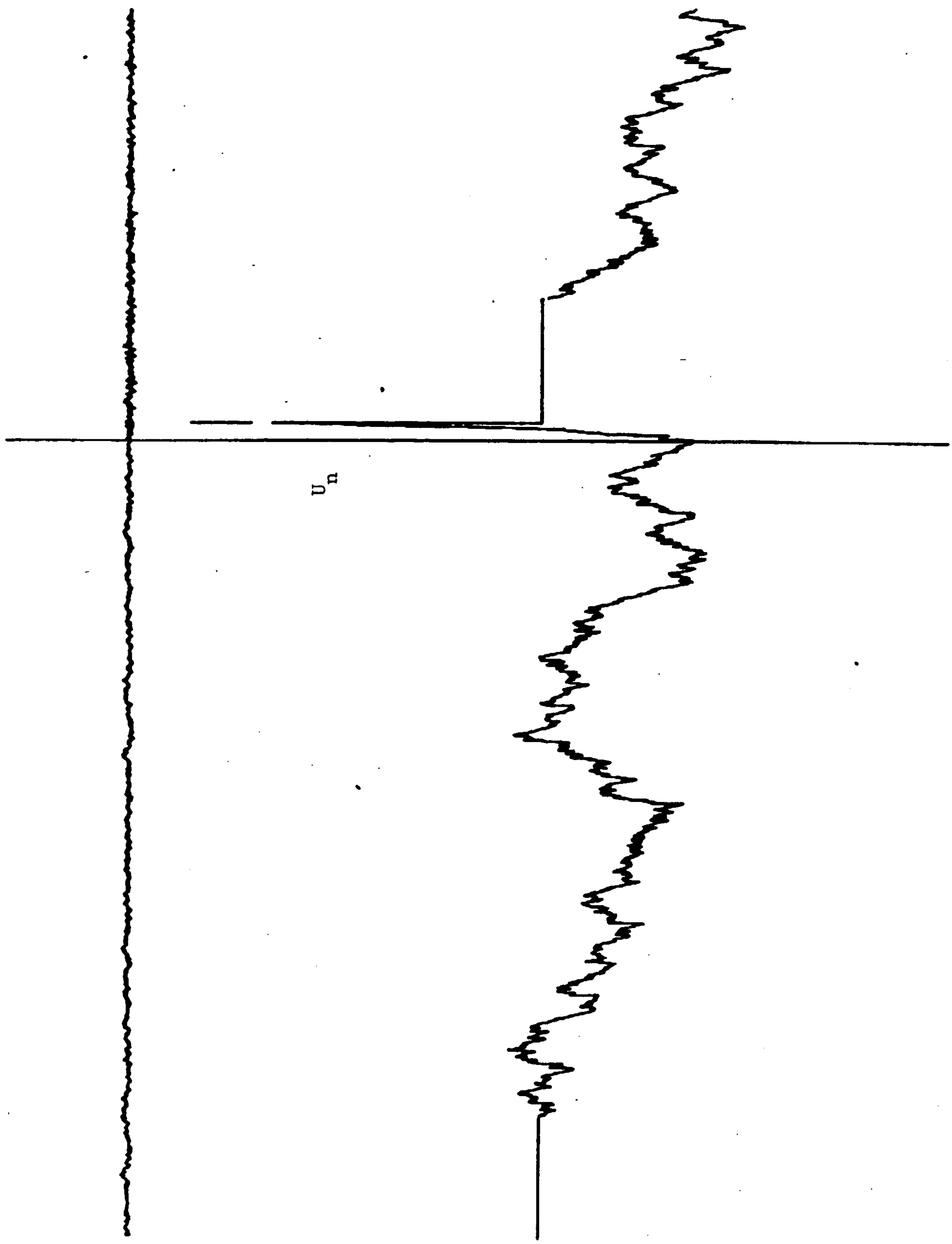


Figure n° 4a

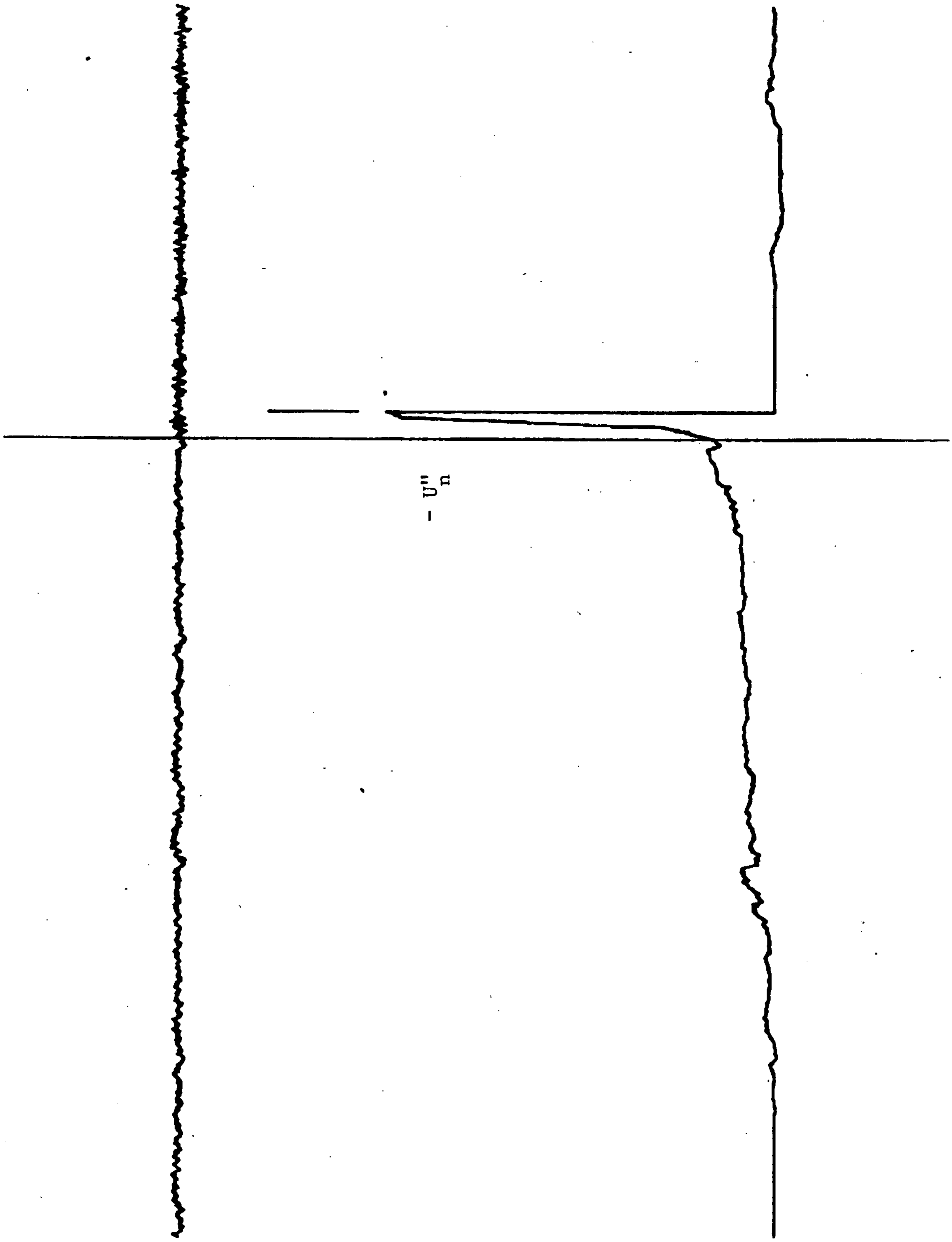


Figure n° 4b

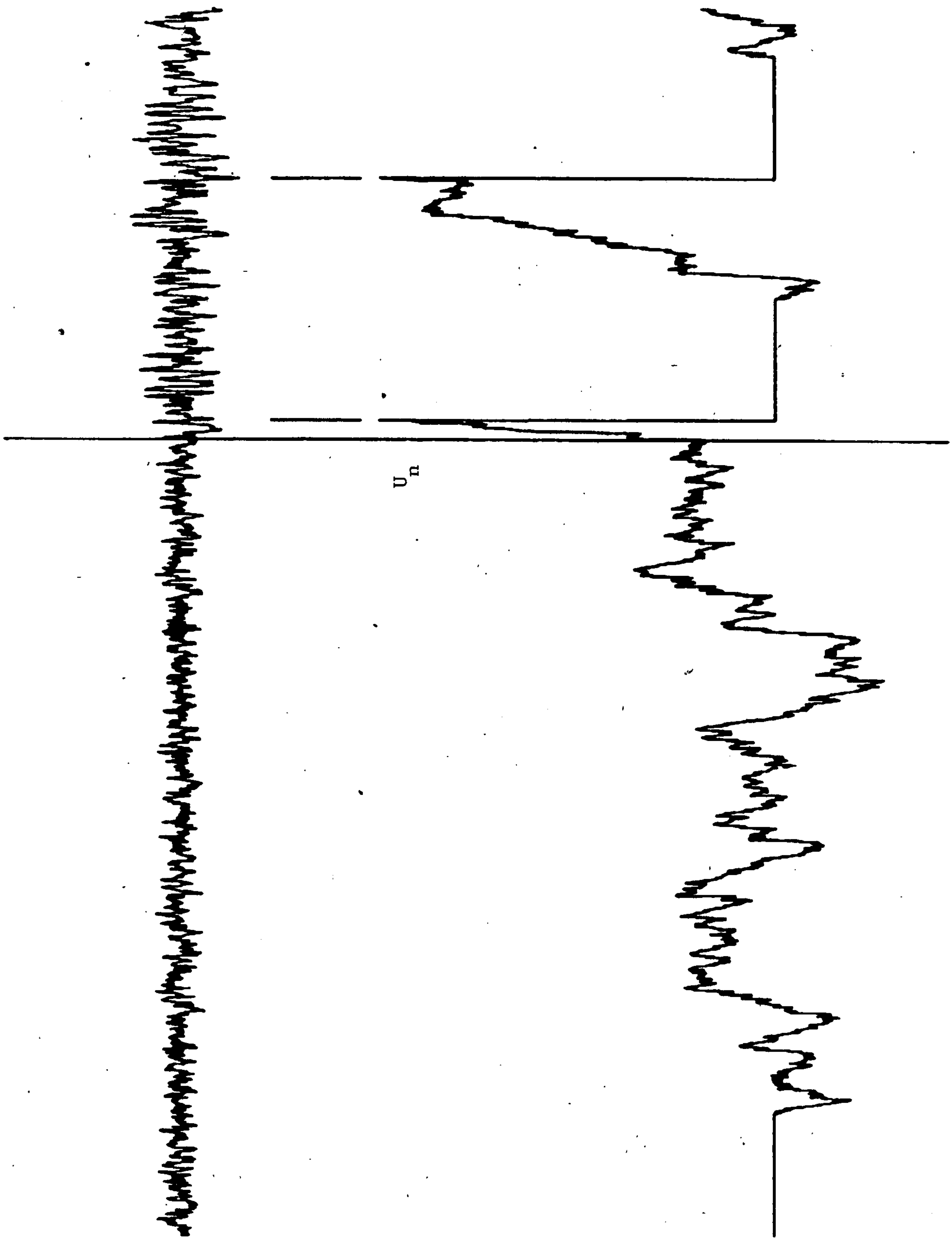


Figure n° 5a

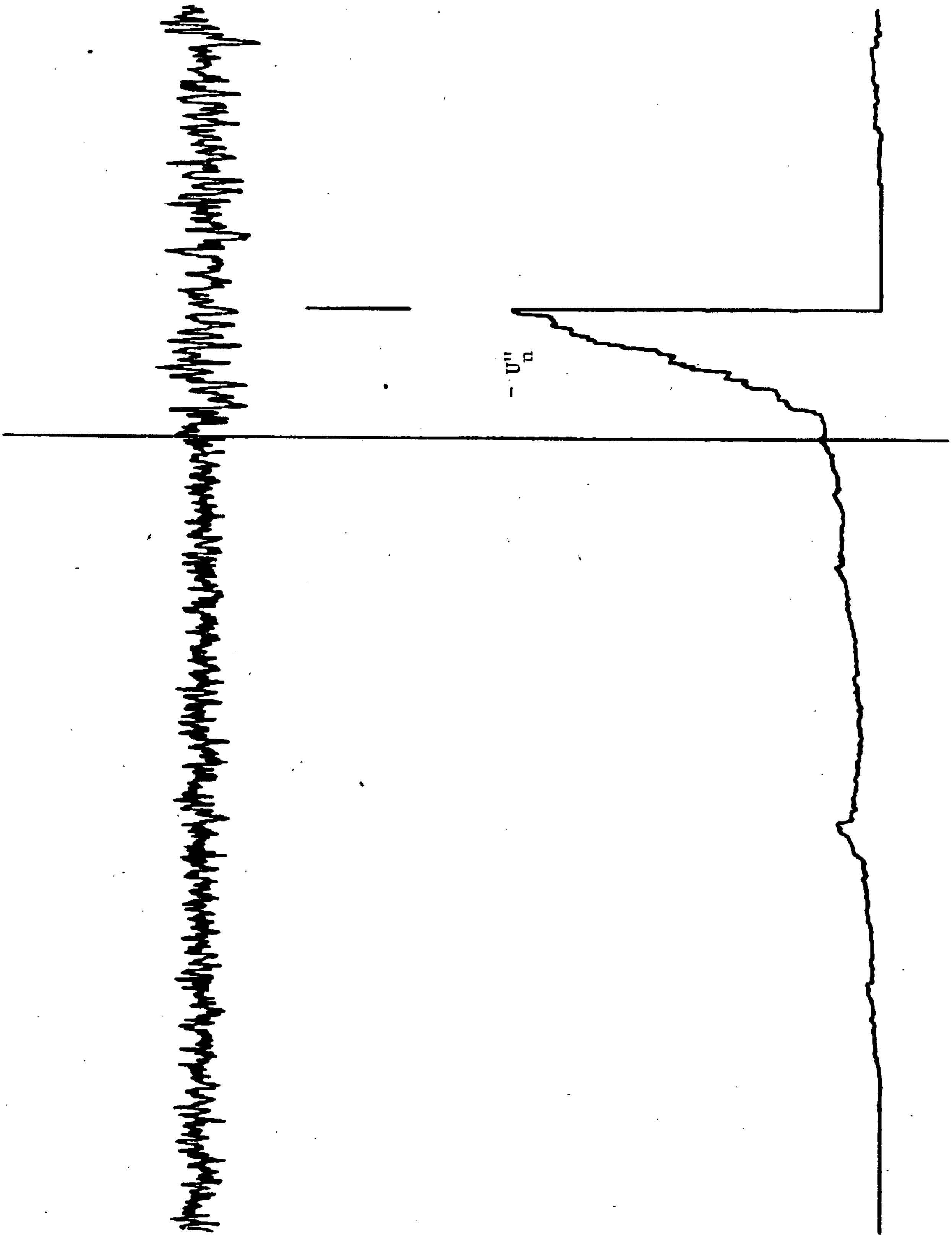


Figure n° 5b

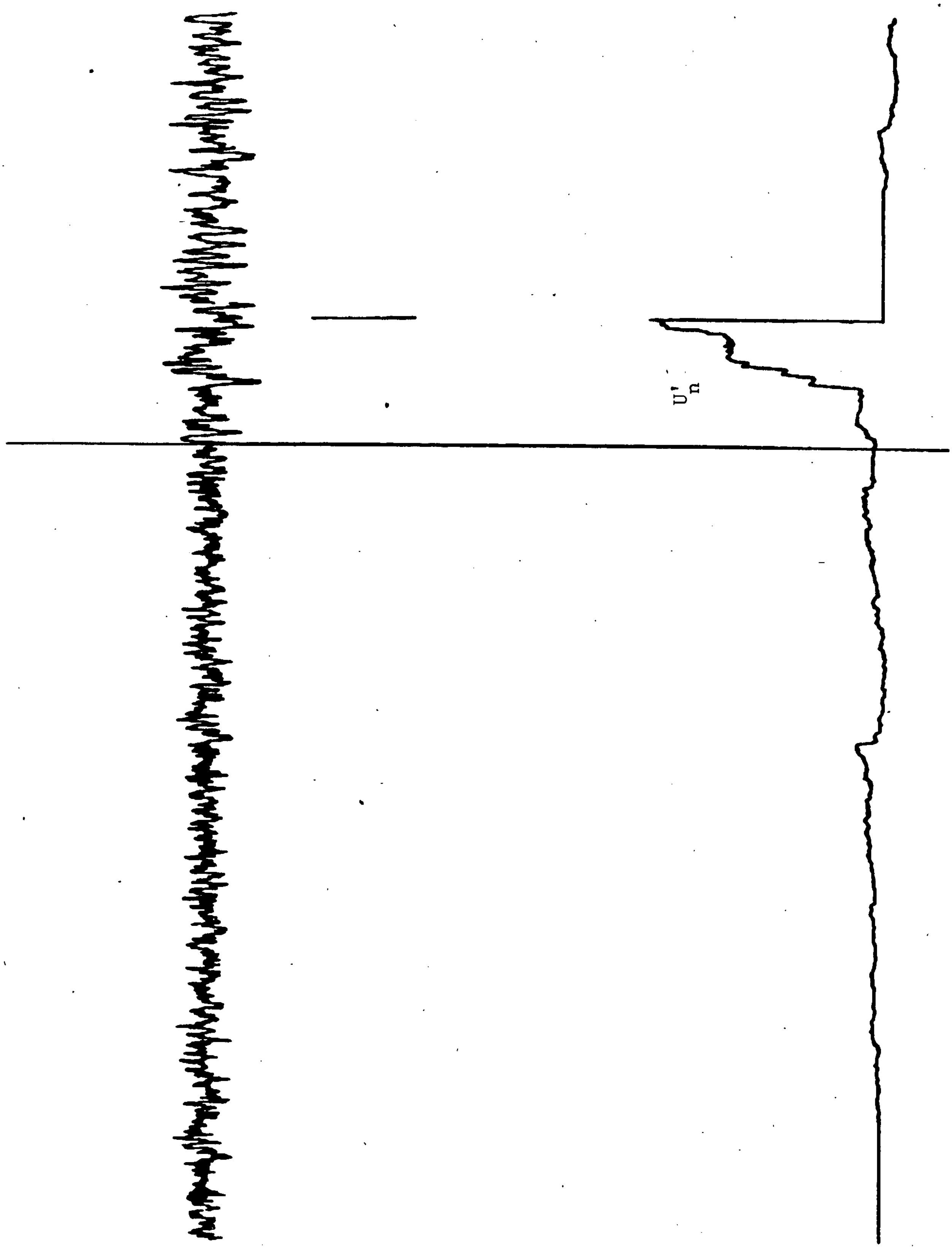


Figure n° 5c

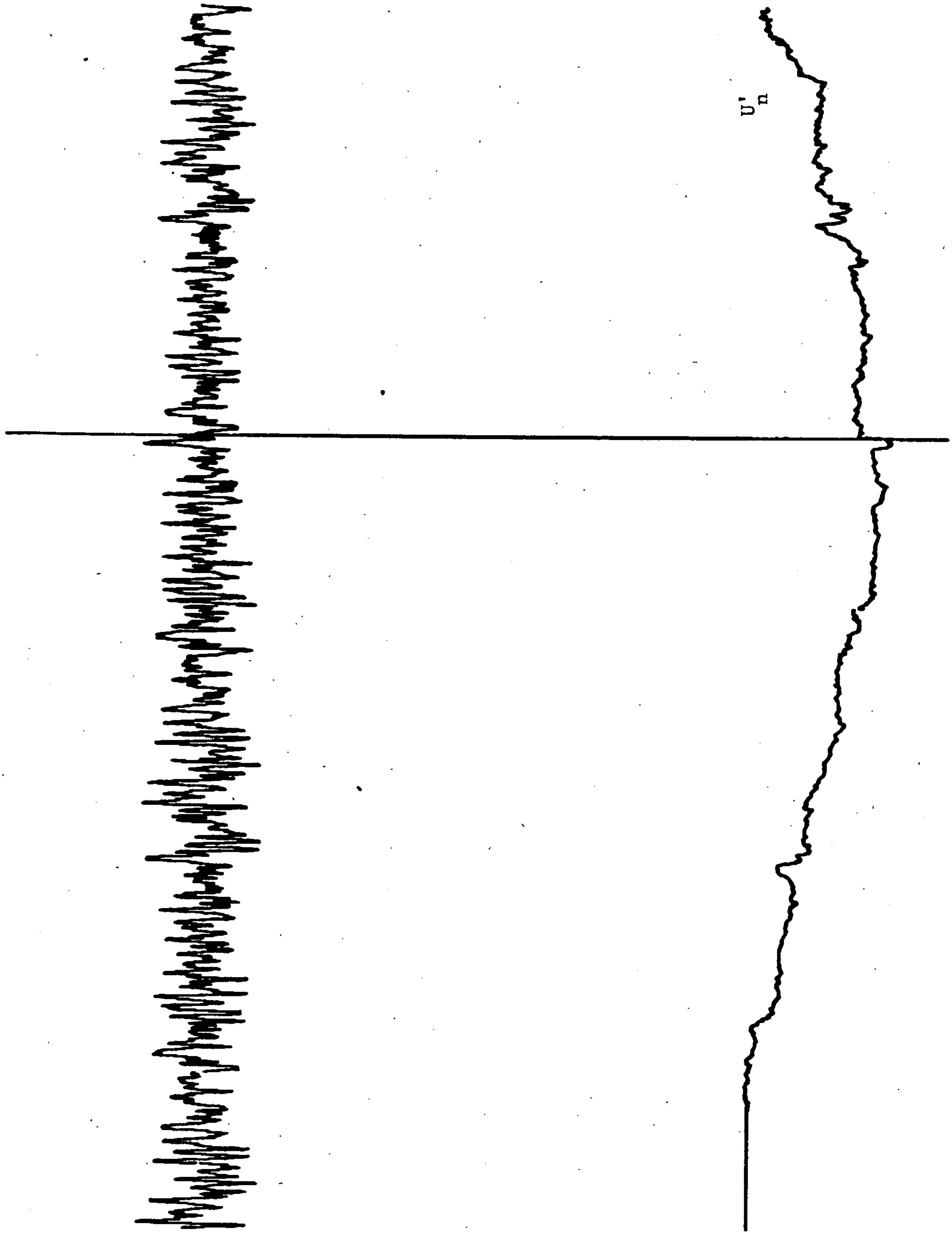


Figure n° 6a

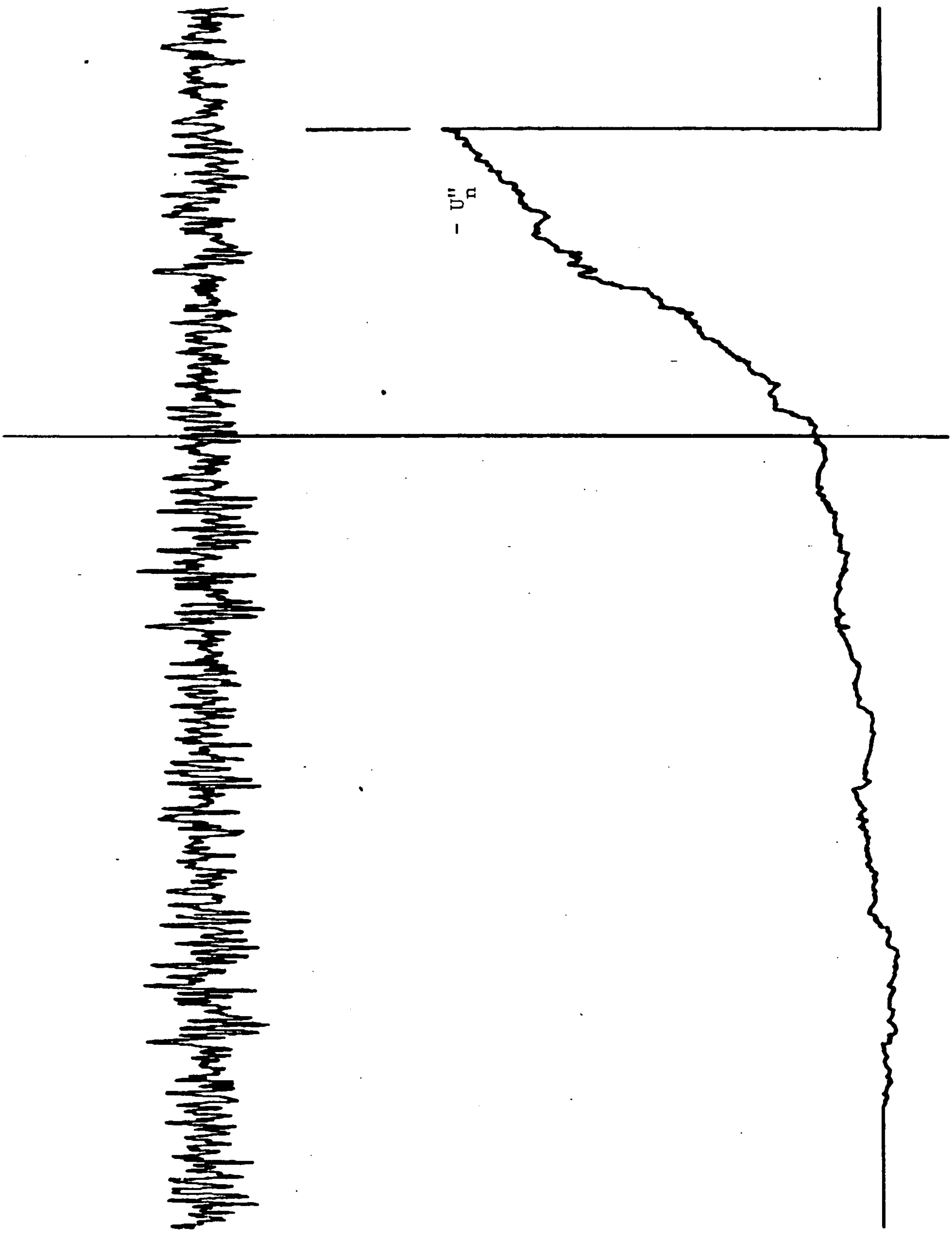


Figure n° 6b

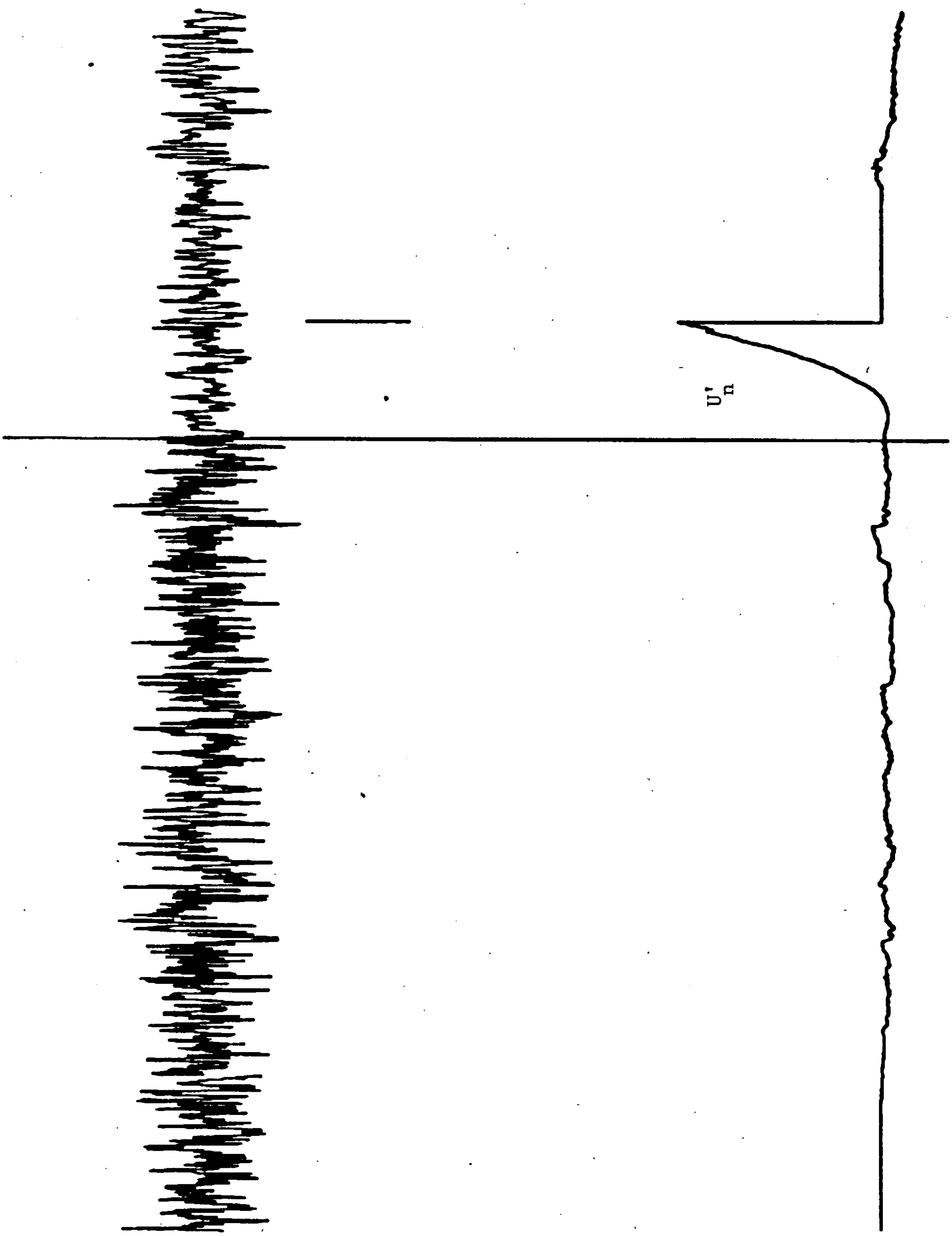


Figure n° 7

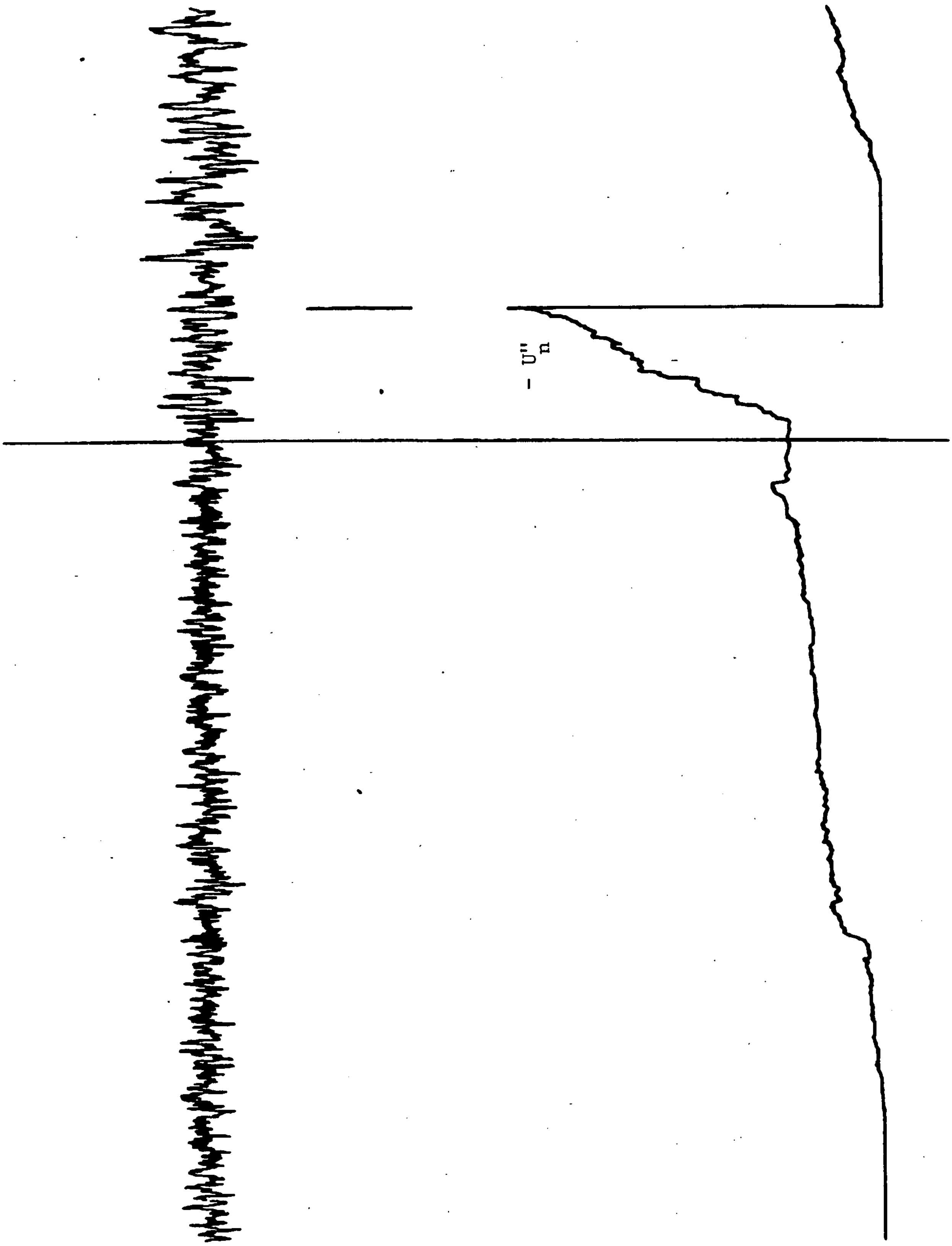


Figure n° 8

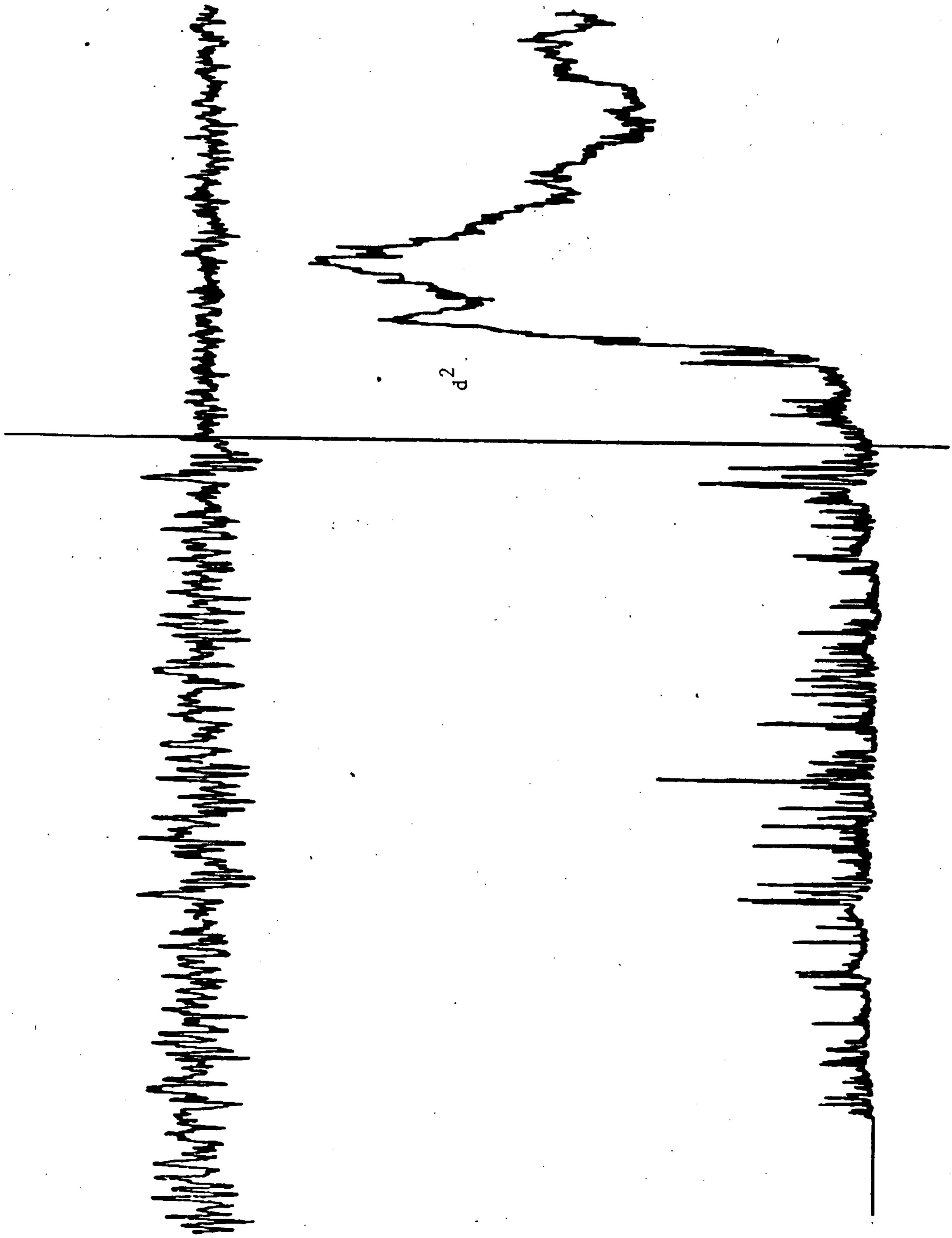


Figure n° 9a

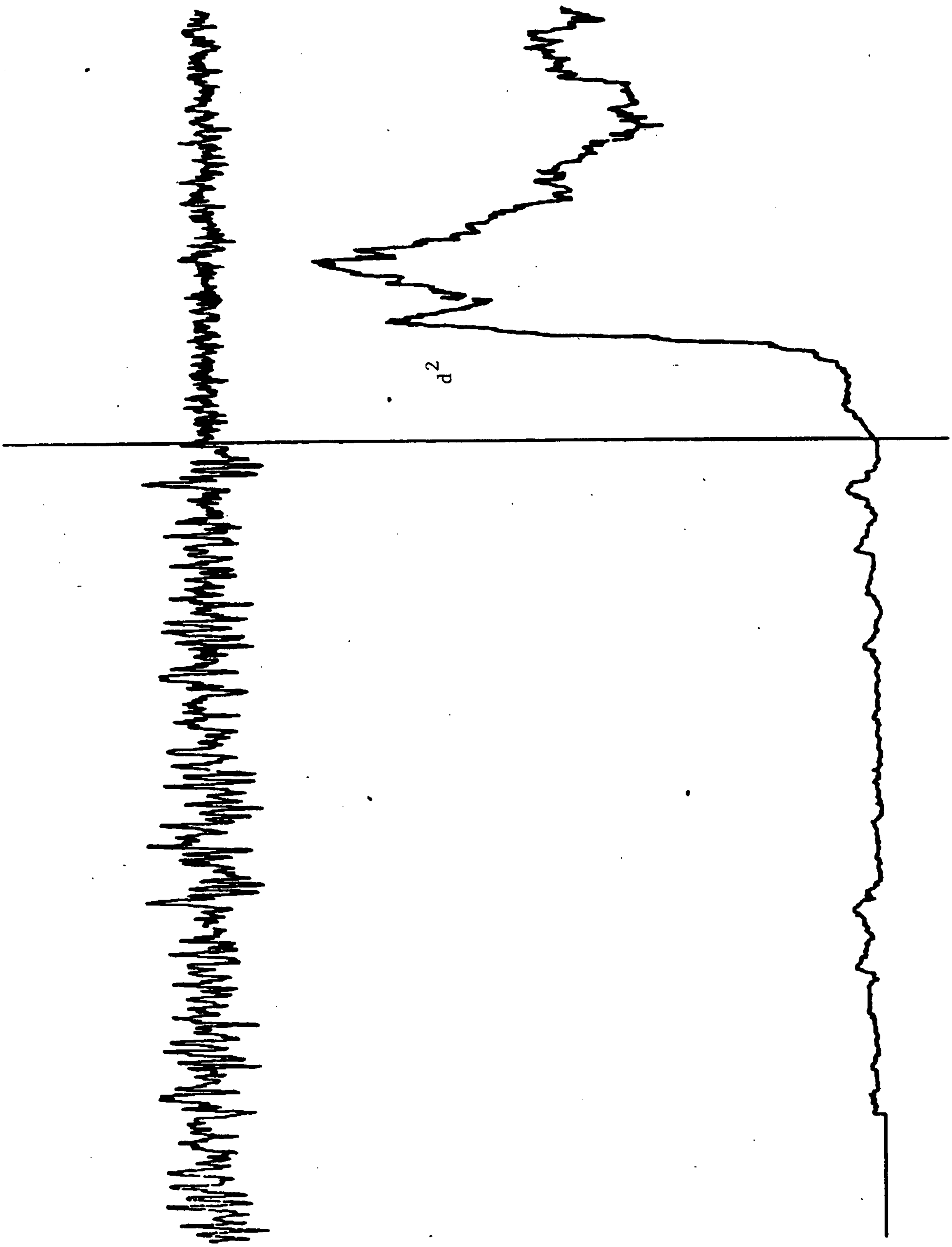


Figure n° 9b

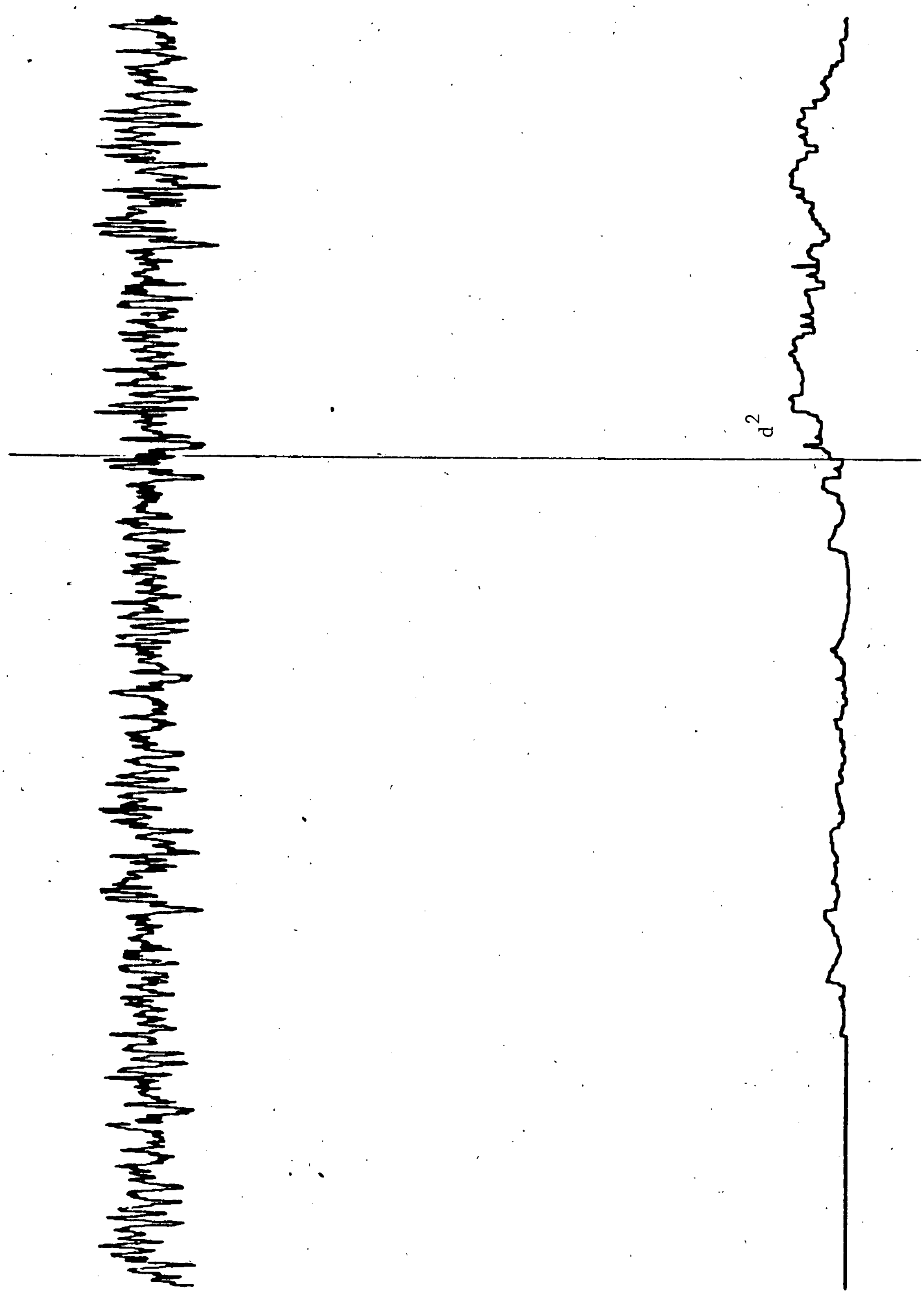


Figure n° 10

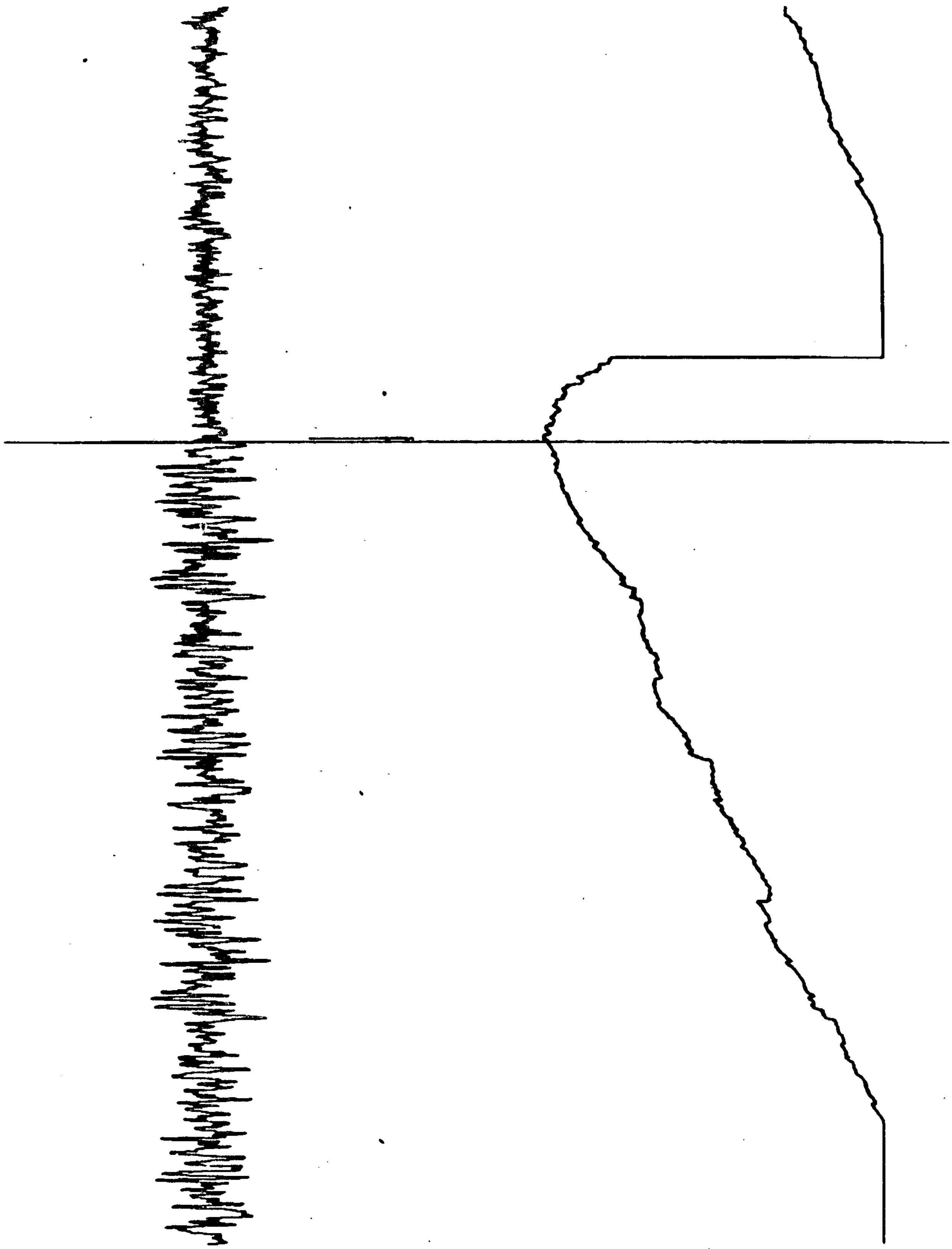


Figure n° 11

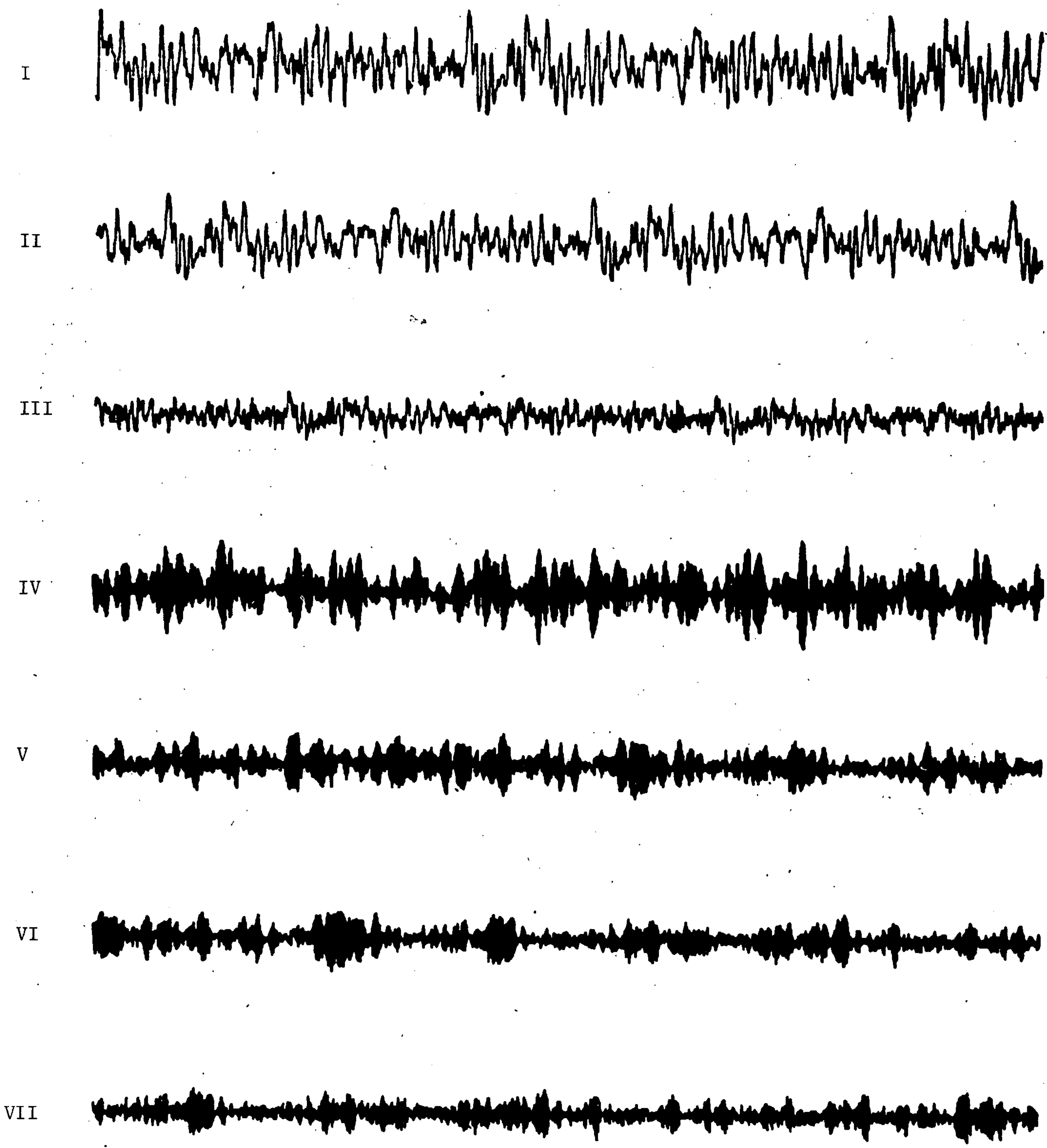


Figure n° 12

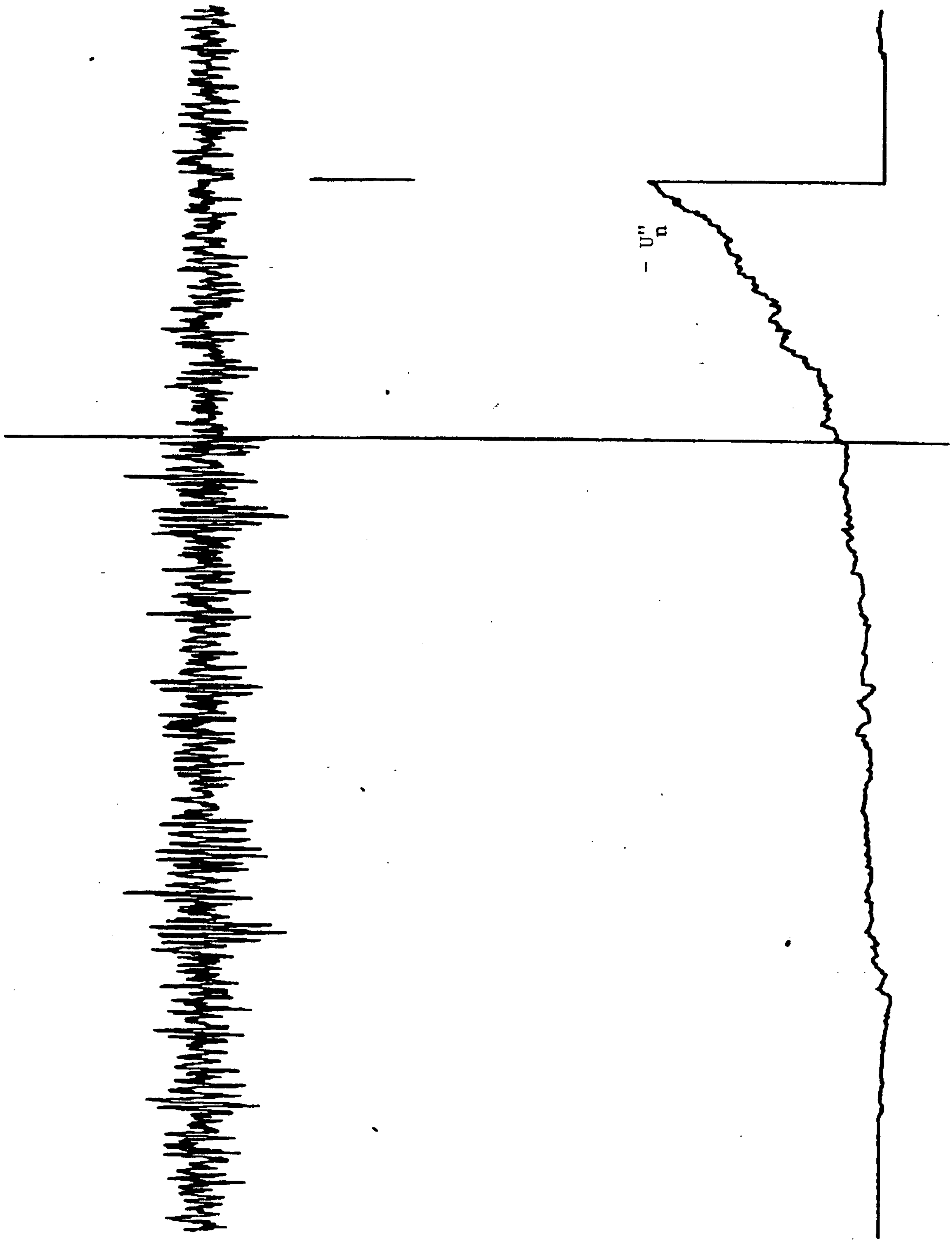


Figure n° 13a

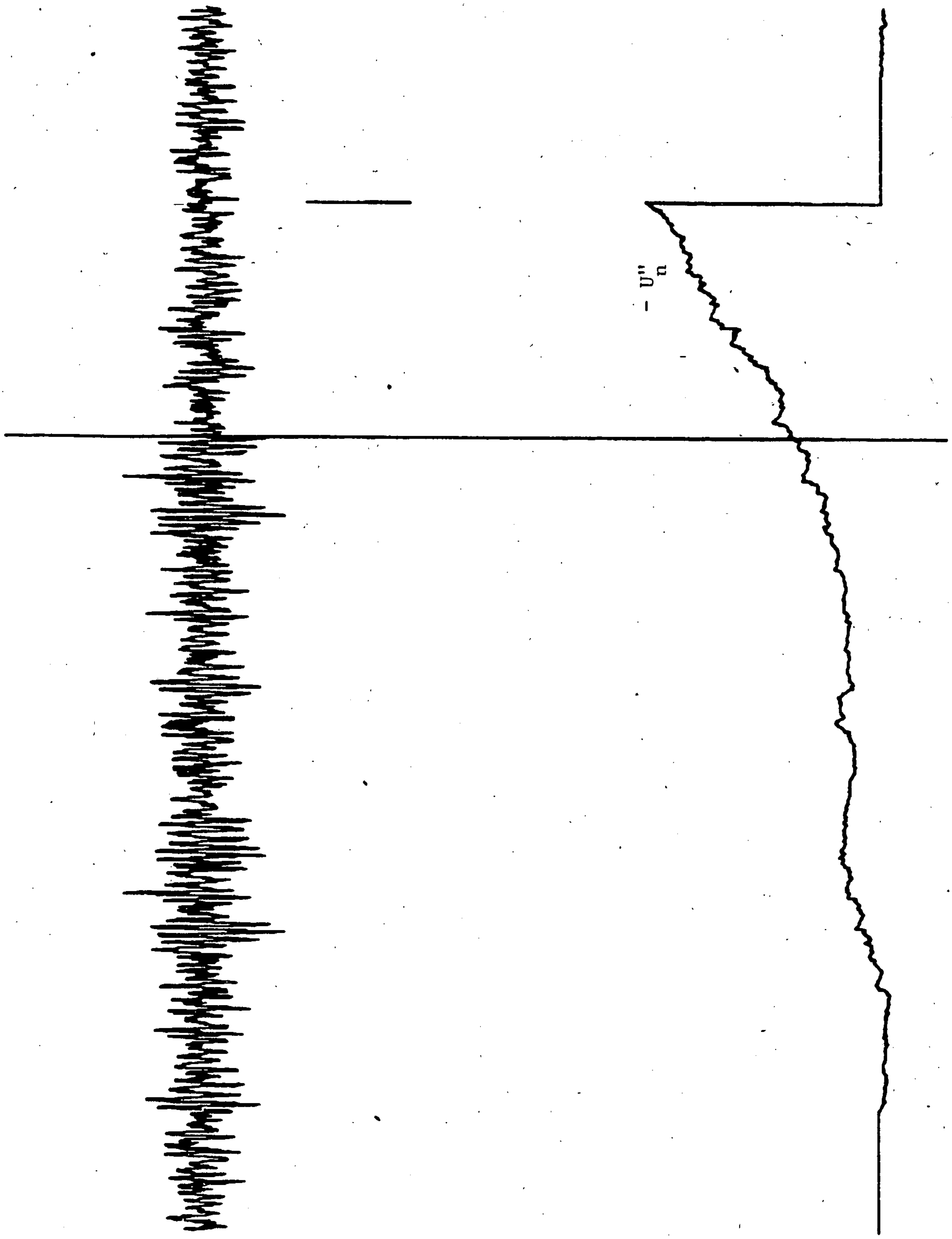


Figure n° 13b

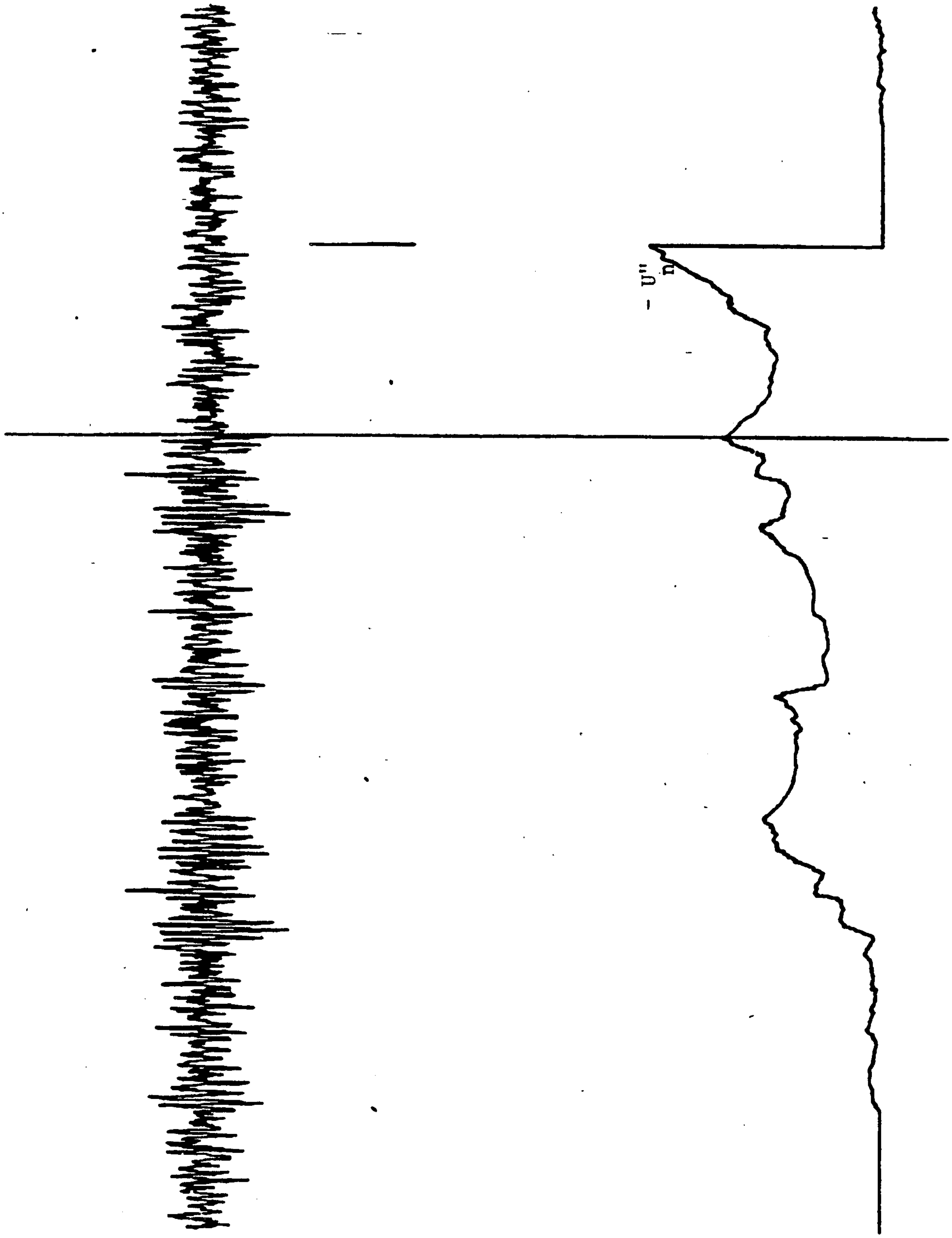


Figure n° 13c

REFERENCES

|1| M. BAGSHAW, R.A. JOHNSON: *"Sequential procedures for detecting parameter changes in a time series model"*. J^al of the American Statistical Association. Vol.72, n°359, pp. 593-597. Sept. 1977.

|2a| M. BASSEVILLE, B. ESPIAU, J. GASNIER : *"Edge detection using sequential methods for change in level. Part I: A sequential edge detection algorithm"*. IEEE Trans. on ASSP. Vol.ASSP-29, n°1, pp. 24-31. Feb. 1981.

|2b| M. BASSEVILLE: *"Edge detection using sequential methods for change in level. Part II: Sequential detection of change in mean"*. IEEE Trans. on A.S.S.P. Vol. ASSP-29, n°1, pp. 32-50. Feb. 1981.

|3| M. BASSEVILLE: *"Survey of statistical failure detection techniques"*. Submitted to IEEE Trans. on A.C.

|4| M. BASSEVILLE, A. BENVENISTE: *"Design and comparative study of some sequential jump detection algorithms for digital signals"*. Submitted to IEEE Trans. on A.S.S.P.

|5| A. BENVENISTE: *"Algorithmes simples d'estimation en treillis pour le problème de la réalisation stochastique"*. Proc. of the Workshop on Fast Algorithms for Linear Dynamical Systems. Aussois, France. Sept. 1981. Ed. by INRIA.

|6| G. BODENSTEIN, H.M. PRAETORIUS: *"Feature extraction from the encephalogram by adaptive segmentation"*. Proc. IEEE. Vol.65, pp. 642-652. 1977.

|7| T. BOHLIN: *" Analysis of EEG signals with changing spectra using a short word Kalman estimator"*. Math. Biosciences. Vol.35, pp. 221-259. 1977.

|8| L.I. BORODKIN, V.V. MOTTL': *"Algorithm for finding the jump times of random process equation parameters"*. Automatica i Telemekhanika. Vol.37, n°6. Part 1; pp. 23-32. June 1976.

|9| B.W. DICKINSON: *"Properties and applications of gaussian autoregressive processes in detection theory"*. IEEE Trans. on I.T. Vol. IT-27, n°3, pp. 343-347. May 1981.

|10| A.H. GRAY, J.D. MARKEL: *"Distance measures for speech processing"*. IEEE Trans. on A.S.S.P. Vol. ASSP-24, n°5, pp. 380-391. Oct. 1976.

|11| R.M. GRAY, A. BUZO, A.H. GRAY, Y. MATSUYAMA: *"Distorsion measures for speech processing"*. IEEE Trans. on A.S.S.P. Vol. ASSP-28, n°4, pp. 367-376. Aug. 1980.

|12| Y. GRENIER: *"Modélisation de signaux, et mesures de distance/distorsion entre signaux: application à la reconnaissance de la parole"*. Proc. of the Workshop on Fast Algorithms for Linear Dynamical Systems. Aussois, France. Sept. 1981. Ed. by INRIA.

|13a| D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER: *"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part I: Identification of Persistent Rhythms"*. IEEE Trans. on B.M.E. Vol. BM-25, n°4, pp. 344-353. July 1978.

|13b| D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER: *"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part II: Identification of Transient Rhythms"*. IEEE Trans. on B.M.E. Vol. BME-25, n°4, pp. 353-361. July 1978.

|14| M. HALL, A.V. OPPENHEIM, A.S. WILLSKY: *"Time-varying parametric modelling of speech"*. Proc. IEEE Conf. on Decision and Control. (New Orléans); pp. 1085-1091. 1977.

|15| D.V. HINKLEY: *"Inference about the change-point from cumulative sum-tests"*. Biometrika. Vol.58, n°3, pp. 509-523. 1971.

|16| A. ISAKSSON, A. WENNBERG, L.H. ZETTERBERG: *"Computer analysis of EEG signals with parametric models"*. Proc. IEEE. Vol.69, n°4, pp. 451-461. April 1981.

|17| N. ISHII, A. IWATA, N. SUZUMURA: *"Segmentation of non-stationary time-series"*. Int. Jal of Systems Sciences. Vol.10, n°8, pp. 883-894. Aug. 1979.

|18| N. ISHII, A. IWATA, N. SUZUMURA: *"Detection of abrupt change and trend in the time-series"*. Int. Jal of Systems Sciences. Vol.11, n°5, pp. 557-566. May 1980.

|19| N. ISHII, H. SUGIMOTO, A. IWATA, N. SUZUMURA: *"Computer classification of the EEG time-series by Kullback information measure"*. Int. Jal of Systems Sciences. Vol.11, n°6, pp. 677-688. June 1980.

|20| F. ITAKURA: *"Minimum prediction residual principle applied to speech recognition"*. IEEE Trans. on A.S.S.P. Vol. ASSP-23, n°1. Feb. 1975.

|21| R.H. JONES, D.H. CROWELL, L.E. KAPUNIAI: *"Change detection model for serially correlated multivariate data"*. Biometrics. Vol.26, n°2, pp. 269-280. June 1970.

|22| D. KAZAKOS, P. PAPANTONI-KAZAKOS: *"Spectral distance measures between gaussian processes"*. IEEE Trans. on A.C. Vol. AC-25, n°5, pp. 950-958. Oct. 1980.

|23| S. LAKEHAL: *"Analyse des signaux aléatoires non-stationnaires par identification récursive"*. Thèse de Docteur-Ingénieur. ENST, Paris. July 1980.

Liste des Publications Internes IRISA

PI 130 **Un système d'enchères simultanées réparti**
M. Banâtre, A.Couvert , 27 pages ; *Janvier 1980*

PI 131 **Méthodes Pseudo-Booléennes**
G. Boulaye , 14 pages ; *Janvier 1980*

PI 132 **Semaine d'étude internationale sur la détection et l'estimation temps-réel des contours et des mouvements dans les images**
Organisation : CCETT, INSA, IRISA (Rennes), , 260 pages ; *Septembre 1979*

PI 133 **Weak solutions for semi-martingales**
J. Pellaumail , 19 pages ; *Mars 1980*

PI 134 **Processus non sequentiel et leurs observations en univers non-centralisé**
Ph. Darondeau , 53 pages ; *Avril 1980*

PI 135 **Synchronisation and protection features for data abstraction**
D. Herman, M.Raynal , 21 pages ; *Juin 1980*

PI 136 **Elaboration et évaluation d'un graphe d'implication pour des données binaires**
I.C. Lerman, R. Gras, H. Rostam , 70 pages ; *Août 1980*

PI 137 **Rapport à la C.E.E. sur les techniques de programmation**
J. André, D. Coan, H. Geist, D. Law, Y. Letertre, H. Schoenen , 96 pages ; *Septembre 1980*

PI 138 **Définition d'un logiciel de transport adapté à la mise en oeuvre d'application burotiques sur des réseaux locaux**
S. Gaucher Cazalis, F. Krier, H. Le Goff , 70 pages ; *Septembre 1980*

PI 139 **Efficacité des algorithmes récursifs en présence de systèmes non-stationnaires.**
A. Benveniste, G. Ruget , 35 pages ; *Août 1980*

PI 140 **Structures de communication extensibles**
P. Le Guernic, M. Raynal , 60 pages ; *Octobre 1980*

PI 141 **Comparaison de tableaux de fréquence**
B. Escoffier , 16 pages ; *Octobre 1980*

PI 142 **Un lemme général de stabilité pour la commande adaptative en déterministe de systèmes non nécessairement à minimum de phase]:**
Cl. Samson , 40 pages ; *Novembre 1980*

PI 143 **Détection, Estimation de l'orientation et saisie d'une cible mobile par proximétrie optique**
B. Espiau , 142 pages ; *Janvier 1981*

PI 144 **Une contribution à l'étude de l'impact de l'informatique sur les organisations**
L. Breton, A. Prod'homme, J. Villard , 58 pages ; *Décembre 1980*

PI 145 **Rupture de modèles statistiques**
M. Basseville, A. Benveniste , 130 pages ; *Mars 1981*

PI 146 **Traitement des questionnaires avec non réponse, analyse des correspondances avec marge modifiée et analyse multicanonique avec contrainte**
B. Escoffier , 38 pages ; *Mars 1981*

PI 147 **Deux files d'attente à capacité limitée en tandem**
J. Pellaumail, J. Boyer , 19 pages ; *Juillet 1981*

PI 148 **Programme de classification hiérarchique : 1) Méthode de la vraisemblance des liens, 2) Méthode de la variance expliquée**
I.C. Lerman , 113 pages ; *Juin 1981*

PI 149 **Convergence des méthodes de commande adaptative en présence de perturbations aléatoires**
J.J. Fuchs , 46 pages ; *Juillet 1981*

PI 150 **Construction automatique et évaluation d'un graphe d'«implication» issu de données binaires, dans le cadre de la didactique des mathématiques**
H. Rostam , 112 pages ; *Juin 1981*

PI 151 **Réalisation d'un outil d'évaluation de mécanismes de détection de pannes]-]Projet Pilote SURF**
B. Decouty, G. Michel, C. Wagner, Y. Crouzet , 59 pages ; *Juillet 1981*

PI 152 **Règle maximale**
J. Pellaumail , 18 pages ; *Septembre 1981*

PI 153 **Corrélation partielle dans le cas « qualitatif »**
I.C. Lerman , 125 pages ; *Octobre 1981*

PI 154 **Stability analysis of adaptively controlled not-necessarily minimum phase systems with disturbances**
Cl. Samson , 40 pages ; *Octobre 1981*

PI 155 **Analyses d'opinions d'instituteurs à l'égard de l'appropriation des nombres naturels par les élèves de cycle préparatoire**
R. Gras , 37 pages ; *Octobre 1981*

PI 156 **Récursion induction principle revisited**
G. Boudol, L. Kott , 49 pages ; *Décembre 1981*

PI 157 **Loi d'une variable aléatoire à valeur R^+** réalisant le minimum des moments d'ordre supérieur à deux lorsque les deux premiers sont fixés
M.Kowalowka, R. Marie , 8 pages ; *Décembre 1981*

PI 158 **Réalisations stochastiques de signaux non stationnaires, et identification sur un seul échantillon**
A. Benveniste J.J. Fuchs , 33 pages ; *Mars 1982*

PI 159 **Méthode d'interprétation d'une classification hiérarchique d'attributs-modalites pour l'«explication» d'une variable ; application à la recherche de seuil critique de la tension artérielle systolique et des indicateurs de risque cardiovasculaire**
B. Tallur , 34 pages ; *Janvier 1982*

PI 160 **Probabilite stationnaire d'un réseau de files d'attente multiclasse à serveur central et à routages dépendant de l'état**
L.M. Le Ny , 18 pages ; *Janvier 1982*

PI 161 **Détection séqentielle de changements brusques des caractéristiques spectrales d'un signal numérique**
M. Basseville, A. Benveniste , pages ; *Mars 1982*

PI 162 **Actes regroupés des Journées de Classification de Toulouse (Mai 1980), et de Nancy (Juin 1981)**
I.C. Lerman , 304 pages ; *Mars 1982*

|24| S. LAKEHAL: *"On-line segmentation of quasi-stationary signals using autoregressive order selection"*. First EUSIPCO. Lausanne (Suisse). Sept. 1980.

|25| D.T.L. LEE, M. MORF: *"A novel innovations-based time-domain pitch detector"*. Proc. 1980 ICASSP.

|26| J.D. MARKEL, A.H. GRAY: *"Linear prediction of speech"*. Springer-Verlag. 1976.

|27| M. MATHIEU: *"Analyse de l'électro-encéphalogramme par prédiction linéaire"*. Thèse de Docteur-Ingénieur. ENST. Paris. Sept. 1976.

|28| I.V. NIKIFOROV: *"Cumulative sums for detection of changes in random process characteristics"*. Automatica i Telemekhanika. Vol.40, n°2. Part 1. pp. 48-58. Feb. 1979.

|29| I.V. NIKIFOROV: *"Modification and analysis of the cumulative sum procedure"*. Automatika i Telemekhanika. Vol.41, n°9; pp. 74-80. Sept. 1980.

|30| D. PERRIOT-MATHONNA: *"Auto-Adaptivité des algorithmes stochastiques récursifs"*. Proc. 8è Colloque GRETSI. Nice, France; pp. 401-408. June 1981.

|31| A.C. SANDERSON, J. SEGEN, E. RICHEY: *"Hierarchical modeling of EEG signals"*. IEEE Trans. on P.A.M.I. Vol. PAMI-2, n°5, pp. 405-414. Sept. 1980.

|32| J. SEGEN, A.C. SANDERSON: *"Detecting changes in a time-series"*. IEEE Trans. on I.T. Vol. IT-26, n°2, pp. 249-255. March 1980.

|33| J. SEGEN: *"Pattern-directed signal analysis: unsupervised model inference, applications to EEG and speech"*. Ph.D. Thesis. Carnegie-Mellon Univ. March 1980.

|34| J.E. SHORE, R.W. JOHNSON: *"Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy"*. IEEE Trans. on I.T. Vol. IT-26, n°1, pp. 26-37. January 1980.

|35| J.E. SHORE: *"Minimum cross-entropy spectral analysis"*. IEEE Trans. on A.S.S.P. Vol. ASSP-29, n°2, pp. 230-237. April 1981.

|36| J.E. SHORE, R.M. GRAY: *"Minimum cross-entropy pattern classification and cluster analysis"*. IEEE Trans. on P.A.M.I. Vol. PAMI-4, n°1, pp. 11-17. January 1982.

|37| J.K. TUGNAIT: *"Detection and estimation for abruptly changing systems"*. Proc. 20th IEEE Conf. on Decision and Control. pp. 1357-1362. 1981.

|38| A.S. WILLSKY : *"A survey of design methods for failure detection in dynamic systems"*. Automatica. Vol.12, pp. 601-611. 1976.

|39a| A.S. WILLSKY, H.L. JONES: "*A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes*". Proc. 13th IEEE Conf. on Decision and Control. pp. 846-853. 1974.

|39b| A.S. WILLSKY, H.L. JONES: "*A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems*". IEEE Trans. on A.C. Vol. AC-21, n°1, pp. 108-112. Feb. 1976.

|40| B. YEGNANARAYANA, D.R. REDDY: "*A distance measure based on the derivative of linear prediction phase spectrum*". Proc. ICASSP Conf. pp. 744-747. 1979.